nature

www.nature.com/nature

Vol 445 | Issue no. 7129 | 15 February 2007

Indian voices

Good policy decisions on science and the environment require sound contributions from official bodies, pressure groups, the media — and scientists themselves.

As India emerges as a global power, one of its greatest assets is its democracy. An important component of democracy, in India as elsewhere, is the thorough public discussion of scientific and environmental issues, as a requisite for the laws, regulations and agencies that will win broad public support and serve national needs.

One of the few points on which most students of politics agree is that the emergence of powerful players who are not aligned to business or to government — commonly known as non-governmental organizations — has broadened and strengthened these debates over the past 20 years or so. Nowhere is this truer than in India.

It is inevitable that some of these voices will be louder than others and in India, no one speaks louder than the Centre for Science and Environment (CSE). The centre, based in New Delhi and founded in 1980 by the journalist Anil Agarwal, has established itself as a significant voice over a wide range of issues (see page 706). Indeed, in many cases, it is the first place Indian media go for a non-official viewpoint on environmental matters.

Stir it up

In some areas, the CSE's work has been laudable and not particularly contentious. It has contributed, for example, to overdue efforts to confront air pollution in urban areas. Sometimes, the centre's role has been more controversial. It played a prominent role in creating a major international fracas over the pesticide levels in Coke and Pepsi in India, for example, chiefly on the basis of comparing the purity of these products with their equivalents in Europe. But some have taken issue with the methodology used, and since these products are bottled in India from local water supplies, others argue that the comparison was unfair.

So although the CSE is an admirably energetic and effective outfit, its perspectives are not universally shared. Some scientists privately complain that the group's influence is out of all proportion to the thoroughness and reliability of its work. There is obviously something in this last contention, as the outfit's public profile is so disproportionate to its tiny scientific staff.

Yet Indian scientists who resent either the CSE's positions or its influence do themselves no favours by carping about either the activities of the Delhi think-tank, or about the media outlets that lap up its output. They should instead look at themselves, and ask if their public influence is commensurate with their own expertise, and with the ever-expanding scope and scale of scientific and environmental policy debates in India.

According to CSE director Sunita Narain, and many journalists, India's scientists too often remain old-fashionedly aloof from the discussions that accompany policymaking. Seeking status and advancement chiefly among their peers, and suspicious of the media's tendency to simplify and exaggerate, scientists who could assist the messy democratic process are inclined, instead, to look down on it. This approach by scientists to science policy is, of course, a global phenomenon. But it is particularly pervasive in India — and particularly inappropriate, given India's vast and pressing need for more public, more thorough, more detailed policy preparation, in areas such as environmental regulation.

"Seeking status and advancement chiefly among their peers, scientists who could assist the democratic process are inclined to look down on it."

Around the world, the scientific community speaks with many voices. In the United States, for example, it has official societies (such as the American Chemical Society), quasi-official leadership (the National Academy of Sciences), unofficial, multi-issue interest groups with large memberships (the Union of Concerned Scientists), voluble individuals (E. O. Wilson), as well as agenda-driven outfits broadly comparable to the CSE (the Center for Science in the Public Interest).

All of them jostle for attention, and all make their voices heard — sometimes even when it matters. In a true democracy, the workings of science and environmental policy more closely resemble an Indian bazaar than a hushed committee room. The sooner Indian scientists join in the fray, the better. ■

Food for thought

Science needs to be better applied to the US food-safety system.

Food-safety oversight in the United States has been in disarray for many years. Responsibility for it is split, on historical grounds, between 15 different agencies in the federal government, operating under at least 30 different statutes. It is past time for Congress to legislate to modernize the entire system.

Late last month, the Government Accountability Office (GAO) added food safety to its list of 'high-risk' federal policies and programmes most in need of reform. The non-partisan GAO is recommending that Congress ask the National Academies to examine new ways of organizing the federal food-safety system. As part of that project, the academies would certainly examine the state of federally funded food-safety research.

"The current fragmented federal system has caused inconsistent oversight, ineffective coordination, and inefficient use of resources," the GAO said in a statement accompanying its updated list, adding

that Congress should "consider a fundamental re-examination of the system".

The alarm sounded by the GAO is likely to resonate with the public, after a three-month stretch last autumn in which outbreaks of food-borne disease killed three people and made more than 500 others sick. The culprits ranged from spinach contaminated with the bacterium *Escherichia coli,* to salmonella-bearing tomatoes, to lettuce that probably infected scores of people with *E. coli* after they ate at Taco Bell and Taco John's restaurants.

These well-publicized incidents stand out against a far larger, latent problem. According to the Centers for Disease Control and Prevention in Atlanta, 76 million Americans contract food-borne illnesses each year, and 5,000 of them die.

A science-backed regulatory system is needed to address an issue on this scale. What has evolved instead over the past century is an irrational and expensive arrangement, whereby officials examine every carcass at every slaughterhouse in the United States every day, but a major food processing plant may escape inspection for a decade.

At the same time, the government research programmes supporting food-safety regulation are neither comprehensive nor coordinated. Research is scattered between often-obscure subdivisions of several departments, from the Environmental Protection Agency to the US Department of Agriculture (USDA). The result is a patchwork of research that, according to a 1998 Institute of Medicine report that reads uncannily like the GAO's latest assessment, "raises serious concerns about duplication of effort and about the linkage of science to attempt to solve food safety problems of the highest priority".

For historical reasons, the USDA is responsible for the safety of meat, poultry and processed egg products, whereas most of the rest — 80% of the food supply — rests with the Food and Drug Administration (FDA).

But the FDA's Center for Food Safety and Applied Nutrition in College Park, Maryland, which bears the primary responsibility for the research that supports food safety at the agency, has seen its budget and staff cut over the past decade. This year, its operating budget will fall to $25 million — a little over half of what it was in 2003. The centre has far less funding for its own science, and next to none for extramural research in areas it needs to learn more about, from microbial ecology to detection methodologies for pathogens in food.

The recent outbreaks caused by bacteria in fresh fruit and vegetables illustrate the paucity of research. Nine years ago, after a string of comparable outbreaks, the FDA issued a set of general recommendations on managing manure, irrigation water and farmworker hygiene to minimize contamination of fruit and vegetables. But it has not done the research to get the data needed to convert these general guidelines into firm, quantified regulations to be implemented on farms, and during food processing and transportation.

"A science-based system is needed. What has evolved instead is an irrational and expensive arrangement."

In the meantime, Democrats in Congress have repeatedly introduced legislation to establish a single US food-safety agency. It is tempting to believe that this approach would produce the necessary coherence in food safety. But past experience of amalgamating parts of the US federal government, from the Department of Energy to the Department of Homeland Security, does not give cause for optimism that such a consolidation would be either efficient or effective.

The best approach may instead involve three more modest steps: an inter-agency panel to properly coordinate food safety; a comprehensive revision of the antiquated and fragmented legislation now governing it, to better reflect today's risks and today's science; and a properly supported, coordinated research programme to inform food safety policy and practice. ■

Methods in full

From now on, *Nature* authors will be able to include more experimental details in their papers.

When in 1960 Theodor Maiman reported the creation of the laser, he did so in about 300 words. Most of these were about the principles. The experiment was described in two sentences (see *Nature* **187,** 493–494; 1960).

Until now, *Nature*'s style of research papers — although more generous in the space allowed than it once was — has been grounded in this telegraphic tradition, allowing comparatively little space for experimental detail. Consequently, with the advent of the Internet, the supplementary material published online has grown voluminous, and nearly ubiquitous — appended to every Article and Letter in this week's issue, for example. And some of it isn't supplementary at all — it is essential for anyone trying to replicate the work.

We have now taken steps to do better justice to what authors have to say, by letting them present full experimental methods as an integral part of their paper. It is clear that more and more people read papers only in their online versions. So we are expanding the online versions of our Articles and Letters, while condensing some of the technical detail in the printed version.

To be specific: in those papers requiring a separate methods section, the online version of the paper will allow authors to include enough detail to satisfy their peers. This is not a 'supplementary' methods file, but a component of the paper, with all the virtues of full-text linking and functionality. It will appear in all online versions, including the authors' versions of papers that can be loaded into PubMed Central and other open-access repositories six months after publication.

But *Nature* also rejoices in being a print publication. We have no wish to leave print readers lacking sufficient understanding of what was done to appreciate the authors' achievements. Accordingly, the print version will include a 300-word summary of the methods. This will also appear in the online version.

Norman Lockyer, the founding editor of *Nature*, might well deplore the loss of brevity in today's scientific reports. But our authors should bear in mind that readers still value succinctness — and that *Nature*'s editors and copy-editors will continue to insist on it. ■

RESEARCH HIGHLIGHTS

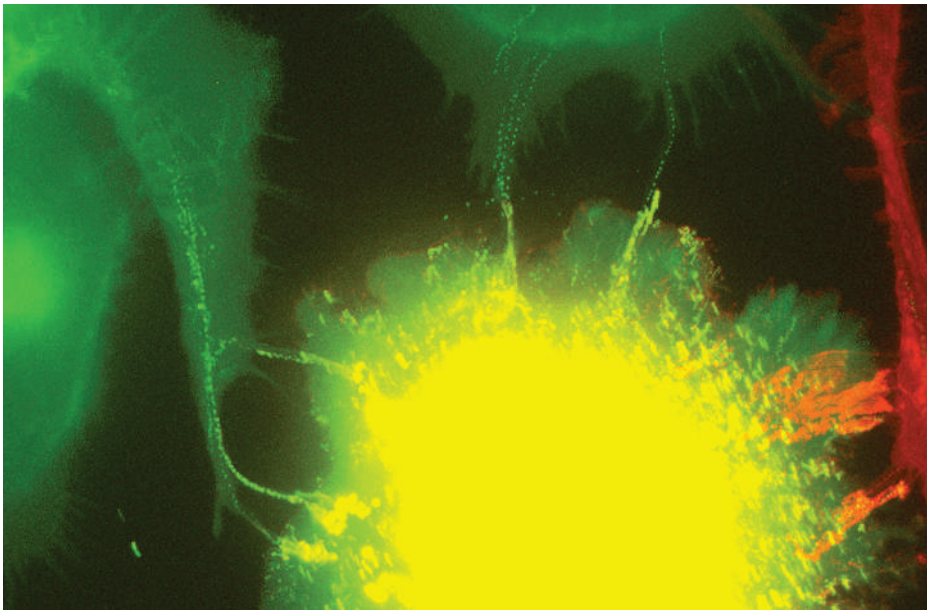
Bridge traffic

Nature Cell Biol. doi:10.1038/ncb1544 (2007)

Videos showing how retroviruses invade cells have revealed that some use an unexpected tactic: they can establish 'bridges' to cross from one cell to an uninfected neighbour.

Walther Mothes of Yale University in New Haven, Connecticut, and his colleagues fluorescently labelled three retroviruses, including HIV, and tracked their movement between cells. Compressing a video of the process into a single picture reveals the tracks of viral particles (which appear green in the image) crossing cytoplasmic bridges. Interactions between a viral envelope protein and proteins on the uninfected cell surface seem to stabilize the bridges.

The experiments were performed in cell culture, but if the findings hold *in vivo* they could suggest new therapies to limit retroviral spread.



W. MOTHES

PHYSICS

Forces of attraction

Phys. Rev. Lett. **98,** 063201 (2007)

For the first time, researchers have measured temperature's influence on the Casimir effect — the attractive force between two objects created by virtual waves that exist in quantum mechanics.

In the 1950s theorists predicted that the Casimir force would show some temperature dependence, but experimentalists have struggled to measure the force between two solid bodies precisely enough to detect the effect. John Obrecht of JILA in Boulder, Colorado, and his colleagues tried a different approach. They measured the force between one solid body — a glass plate — and a small cloud of rubidium atoms held close to the plate. The team saw the force almost treble in strength when the glass plate was heated from room temperature to about 330 °C, in line with expectations.

CELL BIOLOGY

Muscling in

Nature Cell Biol. doi:10/1038/ncb1542 (2007)

Hopes of developing a cell-based therapy for muscular dystrophy have been boosted by work carried out by Giulio Cossu at the Stem Cell Research Institute in Milan, Italy, and his colleagues.

Previously, researchers have turned to muscle-precursor cells known as satellite cells for experimental therapies, but these cells are not able to pass through blood vessel walls, so cannot be conveniently delivered to muscles.

Cossu's team harvested a different type of cell from the blood vessels of human skeletal muscle. The team showed that when these 'pericyte-derived' cells are injected into the arteries of mice with muscular dystrophy, they are incorporated into the muscle and give rise to new, healthy muscle fibres.

MICROFLUIDICS

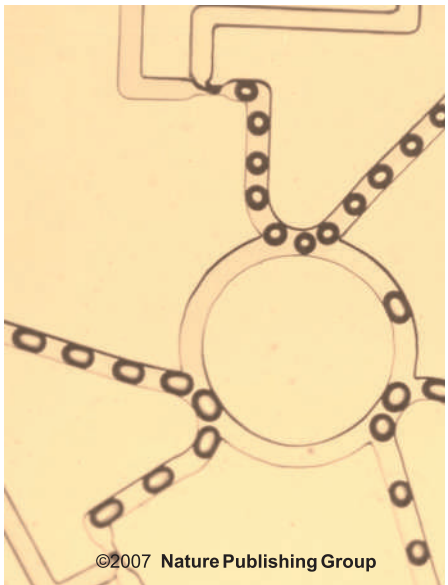
Bubble brain

Science **315,** 832–835 (2007)
Appl. Phys. Lett. **90,** 054107 (2007)
Science doi:10.1126/science.1134514 (2007)

Three independent teams report that they have made simple 'bubble computers'.

Manu Prakash and Neil Gershenfeld at the Massachusetts Institute of Technology in Cambridge, and Levent Yobas and co-workers at the Institute of Microelectronics in Singapore have made logic gates in which binary data are encoded in a fluid flow by the presence or absence of a bubble or an immiscible droplet. The fluid is carried in microscopic channels. At a channel junction, the path a droplet takes depends on the position of droplets in the other channels, making it possible to implement logic operations (the picture, below, shows channels configured as logic gates). Exploiting the same principle, George Whitesides and his colleagues at Harvard University in Cambridge, Massachusetts, designed networks that can encrypt and then decode a signal.

This kind of logic might be used to place microfluidic networks under autonomous control, avoiding the need for externally operated gates and valves.



M. PRAKASH

BACTERIOLOGY

Anchors and ooze

Science **315,** 853–856 (2007)

The 'slime gun' hypothesis put forward to explain how bacteria glide across surfaces has acquired a slime-free rival.

Researchers have previously proposed that bacterial gliders that move without assistance from external structures such as flagella are propelled by the slime they exude. Tâm Mignot and David Zusman of the University of California, Berkeley, and their colleagues suggest that the bacterium might also push against anchor points to move forwards.

The researchers monitored the location of a protein known as AglZ, which is required

for gliding, in *Myxococcus xanthus* cells. The protein formed clusters that remained stationary relative to the surface across which the cells moved. When the cell's lagging end ran into a cluster, the protein dispersed and reorganized at the advancing end. The observations suggest that AglZ defines an anchor point. It's not yet clear whether slime also has a role in this system.

IMMUNOLOGY

Killer speed

Proc. Natl Acad. Sci. USA **104,** 1599–1603 (2007)

Immune cells may sweep up virus-infected cells faster than was thought, according to new research.

Roland Regoes and Rustom Antia of Emory University in Atlanta, Georgia, and their colleagues analysed data from experiments in mice, which measured how quickly virus-laden cells were cleared from the animals' spleens. The experiments compared the response of mice relying on effector killer T cells, produced on exposure to the virus, to the response of those equipped with memory killer T cells, which are associated with long-term immunity.

The researchers found that the two types of killer T cell are equally efficient at reducing the number of infected cells, and calculated a 'rate constant' that expresses how quickly they do so. This constant was up to 100-fold higher than previous studies had suggested — a discrepancy that the team says will need further work to resolve.

CANCER BIOLOGY

Guardian gene

Cell **128,** 459–475 (2007)

Researchers in the United States have teased out the secret of a chromosomal region that is deleted in many human cancers.

Alea Mills from Cold Spring Harbor Laboratory, New York, and her colleagues have identified within the region, known as *1p36*, a gene that protects against cancer. Scientists have for decades assumed that the region harbours such a tumour suppressor, but until now its identity had eluded them.

Mills and her co-workers pinned down the gene by engineering mice to express different numbers of copies of the region corresponding to human *1p36*, then blocking the genes that it contains one by one.

The team thinks that the tumour-suppressor gene, known as *CHD5*, may act as a sort of master switch, remodelling the chromatin that packages DNA to control the cell's complex of tumour-suppressing networks.

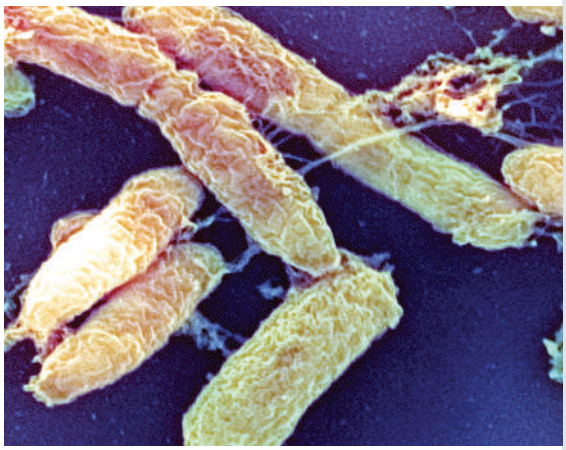
MICROBIOLOGY

Plague aided by phage

Mol. Microbiol. **63,** 1145–1157 (2007)

The bacterium responsible for plague (*Yersinia pestis*, pictured below) ravaged medieval Europe and is resurgent in areas of Africa and Asia. Anne Derbise of the Pasteur Institute in Paris, France, and her colleagues present genetic evidence that helps to explain the microbe's virulence.

The genome of *Y. pestis* has previously been shown to incorporate a sequence encoding an elongated virus, known as a filamentous phage. The presence of such phages in other pathogens is associated with virulence. The researchers confirm that the phage plays a part in the pathogenicity of *Y. pestis* — showing that it is absent from the genome of a less virulent ancestor, and that removal of the phage DNA makes *Y. pestis* less infective.



G. GAUGLER/SPL

MATERIALS SCIENCE

How memories melt away

Phys. Rev. Lett. **98,** 055505 (2007)

Commercial rewriteable DVDs and prototype memory devices record information in phase-change materials such as germanium antimony telluride ($Ge_2Sb_2Te_5$ or GST) that switch rapidly between crystalline and amorphous states when heated with a laser. The speed of the phase change, which can occur in nanoseconds, is crucial to the perfomance of such devices, but no one knows quite how it happens.

Zhimei Sun of Uppsala University in Sweden and co-workers say that the melting of GST is decidedly odd. Their computer simulations suggest that the atoms form chains and tangled clusters as the crystals melt in two dimensions, whereas in the third dimension the atoms remain in orderly layers. They believe this accounts for the material's quick phase change.

JOURNAL CLUB

Galina Khitrova
University of Arizona, Tucson, USA

An expert on instabilities jumps from optically bound plastic beads to the brain.

It's not often that reading scientific papers turns my mind to the melancholic work of great Russian writers, but a recent one did.

The paper reports observations of 'bistability' in a simple optical system. Bistable systems have two stable output states for the same input. In this case, the researchers had studied the behaviour of two plastic spheres, trapped side-by-side in a pair of counter propagating laser beams (N. K. Metzger *et al. Phys. Rev. Lett.* **96,** 068102; 2006). They found that the beads could adopt two stable arrangements, differing in the beads' separation.

Bistability arises in optical systems that show nonlinear responses to changes in light intensity and include some kind of feedback process. Here, one bead feels the position of the other because each affects the light field around it, creating the necessary feedback.

The researchers modelled how the two stable states come about, combining equations that describe the propagation of the light with others that predict the forces on the beads. I was impressed by how many physical effects are taken into account in the model.

And this is what turned my thoughts away from the physics of my research to the literature of my homeland. It is believed that some Russian authors, including Leo Tolstoy, may have suffered from what is now known as a bipolar disorder, characterized by states of euphoria and depression.

I have wondered before whether bistability in optical systems might serve as a simple model to help understand the mechanisms that underlie bistability in the human brain. Papers such as this one put that challenge in perspective — modelling a system that involves just two beads is already nontrivial.

NEWS

Broad sweep of genome zeroes in on diabetes

In 1918, Ronald Aylmer Fisher, an evolutionary biologist and pioneer of modern statistics, published a paper on the genetic causes of disease that brought together two rival factions. Geneticists promoted a paradigm in which diseases worked a lot like Mendel's pea plants, with just one or two genes responsible for each condition. Biometricians, however, advocated a continuous distribution of phenotypes. Fisher suggested that many mendelian traits could result in the continuous distribution of a disease. In doing so, he established the conceptual basis for the search for complex disease genes that continues today.

But Fisher's theories had a more immediate impact on animals and agriculture than on medicine — in people, it's much easier to study and measure mendelian diseases and traits. Even the much-heralded Human Genome Project in the 1990s didn't help as much as expected. The two methods traditionally used to hunt down disease genes are linkage analysis, which uses large family trees to work out which genes are shared by affected individuals, and the candidate-gene approach, which uses physiological clues to narrow down potential culprits. But when it comes to complex conditions such as heart disease or diabetes, in which multiple environmental and genetic factors combine, neither method is very powerful. Scientists have identified just a handful of disease genes, along with lots of weak, unconfirmed hits.

Now, after a shaky start, hopes are high that a more ambitious breed of genetics study can finally crack the problem. Modern gene-chip technology combined with recently published maps of human genetic variants — particularly the 'HapMap' that groups together related variants called single nucleotide polymorphisms — now enables the entire genomes of thousands of people to be scanned. Many population geneticists and disease researchers think that such genome-wide association (GWA) studies will identify genes that confer even a small extra risk of disease.

It has taken time for big GWA studies to be completed. "Many people didn't know how much association studies would deliver," says Peter Donnelly, a lead investigator of the Wellcome Trust Case Control Consortium, which began collecting samples for GWA studies in 2005.

Yet new results, including a study on type 2 diabetes published this week (R. Sladek *et al.* *Nature* doi:10.1038/nature05616; 2007) suggest that the GWA approach will bear fruit, and lots of it. The group, led by Constantin Polychronakos of McGill University in Quebec, Canada, studied some 393,000 single nucleotide polymorphisms in the genomes of around 700 patients with type 2 diabetes, and 600 controls; the findings were then confirmed in another 14,000 people. The researchers identified four genomic regions that confer a significant risk

Berkeley's energy deal with BP sparks unease

SAN DIEGO

Debate is intensifying at the University of California, Berkeley, over a $500-million energy research partnership with BP. The energy company announced on 1 February that it will fund a decade of alternative-energy research by Berkeley and its partners. But concerns are spreading across the campus about the propriety of the industrial relationship.

Some fear that the pact — for which final details are still being worked out — could be a repeat of a controversial $25-million contract that the university entered into in 1998 with the biotech giant Novartis (see *Nature* 399, 5; 1999). That deal expired in 2003, amid criticism that the academic freedom of some university researchers had been compromised (see *Nature* 426, 591–594; 2003).

Berkeley and its partners, the Lawrence Berkeley National Laboratory (LBNL) and the University of Illinois at Urbana-Champaign, beat four other universities in a six-week competition for cash from the energy company. The research agenda for the initiative has been drawn up by a group of Berkeley researchers. It aims to use biotechnology to develop new energy sources, for example genetically modifying plants for use as fuels, and using enzymes to convert plant material into fuel more efficiently, as well as studying how these new methods might affect agriculture and society.

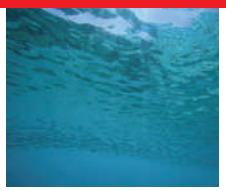
The University of California, Berkeley, is to be home to an energy research centre in a partnership with BP.

M. E. GIBSON/CORBIS

During forthcoming budget negotiations, California's governor Arnold Schwarzenegger plans to push the state for $40 million in bonds to pay for a new building called the Energy Biosciences Institute, where BP-funded researchers would work. The building would house university professors and students, along with perhaps 50 industry scientists.

Industry funds a lot of research on public and private university campuses, and it's fairly common for companies to have labs located near institutes where industry and academic researchers work together — as Intel and Yahoo do at Berkeley, for example. But it's rare

The idea of hunting across the whole genome for links to disease is beginning to pay off.



CARBON GOES DEEP
Studies show CO_2 has reached the bottom of the ocean.
www.nature.com/news

BUZZ PICTURES/ALAMY

P. MENZEL/SPL

to developing the disease. Along with the previously identified *TCF7L2* gene, these regions together might account for 70% of the genetic risk for the disease.

Of the four new genes, Polychronakos says that the best hit is *SLC30A8*, a zinc transporter, which is important because zinc assists with the packaging and secretion of insulin. The importance of the other hits, which have roles in pancreas development and insulin degradation, are less clear, he says.

In general, Polychronakos believes the most likely result of GWA findings will be diagnostic tests that predict who is at high risk of disease. He also envisions using genotypes to determine who would respond to which drugs, in the much-anticipated era of personalized medicine. Drug leads based on gene finds are less likely, Polychronakos thinks. Often, the disease genes uncovered are transcription factors, which he says make poor drug targets. But he does suggest that in the case of the diabetes study the zinc transporter could make a future drug target, or zinc could be used in treatment.

"Many people didn't know how much association studies would deliver."

Geneticist David Altshuler of the Broad Institute in Cambridge, Massachusetts, is more excited about prospects for new therapies. He cites the case of cholesterol, in which a study of heart disease uncovered gene mutations related to cholesterol, allowing researchers to develop a group of drugs called statins.

Either way, the diabetes paper promises to be the first of several big finds. Donnelly says that in the next six months or so, the Wellcome Trust Case Control Consortium plans to publish genes associated with seven complex diseases, including coronary heart disease and rheumatoid arthritis. The formation of large collaborations focused on particular diseases — such as the FUSION study for type 2 diabetes — should help too, allowing researchers to share massive sets of genotype samples.

That's not to say that there aren't challenges ahead. If a gene is particularly rare, or if a disease involves dozens of genes that each have a small effect, then even sample sizes of several thousand might not pick up the signal. Donnelly, though, is more optimistic about the promise of the technique than he was three years ago. "The way I think about it is that some diseases will need much larger studies for us to be convinced of an effect," he says.

Diseases that have a wide variety of symptoms and physiological characteristics, such as schizophrenia, may be more difficult to address. "My advice: find as homogeneous a phenotype as possible," says Polychronakos. For example, he and his colleagues excluded obese people from their study so that they could focus on diabetes genes that confer risk independently of obesity.

Still, based on the number of papers coming up in 2007, Altshuler expects a major jump in the number of solid leads for disease genes, something neither linkage analysis nor the candidate-gene approach could match. Modern biology may finally have begun to bring technological and scientific rigour to Fisher's decades-old insights. ■

Gene Russo

for industry to house its scientists in public buildings on state university property.

To seasoned industry critics, such as Berkeley entomologist Miguel Altieri, the deal is just one more step in "the rapid, unchecked and unprecedented global corporate alignment of the world's largest agribusiness, biotech, petroleum and automotive industries". He fears that for "a relatively small investment", BP can benefit from public resources and cash in on inventions developed with taxpayers' money.

Berkeley's leaders have taken steps to address expected faculty concerns about the deal's impact on campus freedom and intellectual property. Berkeley executives consulted leaders in the Academic Senate early in the bidding process, and the feedback was blunt, says senate chair William Drummond: "No more Novartis deals."

Drummond says that a senate representative will be closely involved in all negotiations with BP: "We want to make sure the university's pockets are not picked." Beth Burnside, Berkeley's vice-chancellor for research, who is helping to negotiate the deal, says that established campus controls will monitor conflicts of interest and issues of intellectual property and academic freedom. BP may have first right of refusal on Berkeley inventions from the collaboration, she says, but licences won't necessarily be exclusive.

The new institute will be built conveniently close to another energy-research endeavour, the Helios Project. This will be directed by officials at LBNL, which is a Department of Energy facility managed by the University of California. Helios has been in the pipeline for three years, as part of a shift in LBNL's mission towards renewable-energy projects, but has been unable to secure start-up funding from the Department of Energy. Instead, Schwarzenegger will push the state legislature for $30 million in bonds to pay for the building.

More controversially, the BP competition occurred alongside a volatile political campaign in California to create a $4-billion public research programme into alternative energy sources, funded via a severance tax on oil firms. Energy companies spent $108 million on advertisements against the measure, Proposition 87, on last November's ballot. Schwarzenegger refused to back Proposition 87, and critics are upset that, instead, he is supporting a deal that they see as enabling one of those energy companies to benefit from public facilities. Schwarzenegger argues that the BP deal fits California's plans for developing cleaner energy in an economical manner.

The losing bidders were the Massachusetts Institute of Technology; the University of California, San Diego; Imperial College London and the University of Cambridge, UK. Imperial's rector Richard Sykes notes that his university had costed its bid so no public funds would be used. He says BP told Imperial that its bid wasn't economical. "We thought that was interesting," he comments.

Scrutiny of the Energy Biosciences Institute and its mission is likely to continue. "The debate is going to increase," says Berkeley physicist Daniel Kammen, who helped write Proposition 87 and the BP proposal. "But I think this is worth trying. BP knows this is a difficult gig; they are not shying away from it." ■

Rex Dalton

Disputed inquiry clears bubble-fusion engineer

An inquiry has exonerated nuclear engineer Rusi Taleyarkhan of misconduct with respect to allegations made internally at Purdue University in West Lafayette, Indiana, officials announced last week. But the announcement may raise more questions than it answers: researchers in the field have criticized the university for failing to say whether the inquiry considered their concerns that the work may be fraudulent.

Taleyarkhan claims to be able to produce fusion by collapsing bubbles in deuterated liquids. His work promised to improve prospects for developing a clean source of energy, but independent scientists have not been able to replicate the result. The work had been subject to several internal allegations of misconduct, including the fact that Taleyarkhan cited a paper by his student and postdoc as "independent" confirmation of his findings[1,2].

Purdue announced on 7 February that "the committee determined that the evidence does not support the allegations of research misconduct and that no further investigation of the allegations is warranted". It has refused to specify the content of the allegations that it considered, except to say that they were "internal".

Institutional proceedings involving Taleyarkhan began in March 2006, after concerns about his work were reported by *Nature*[3]. Purdue's provost, Sally Mason, responded by saying that the university would undertake an objective review. In June 2006, the university said that the review was complete, but declined to make its findings public. Last week's announcement referred to the findings of a second internal inquiry subsequently appointed by Purdue's dean of engineering, Leah Jamieson.

Taleyarkhan has told several news outlets that he feels "vindicated". But critics have questioned the validity of Purdue's proceedings, and in particular, the apparent decision to limit its inquiry to internal allegations, yet possibly ignoring the concerns, including fraud, communicated by external researchers in the field.

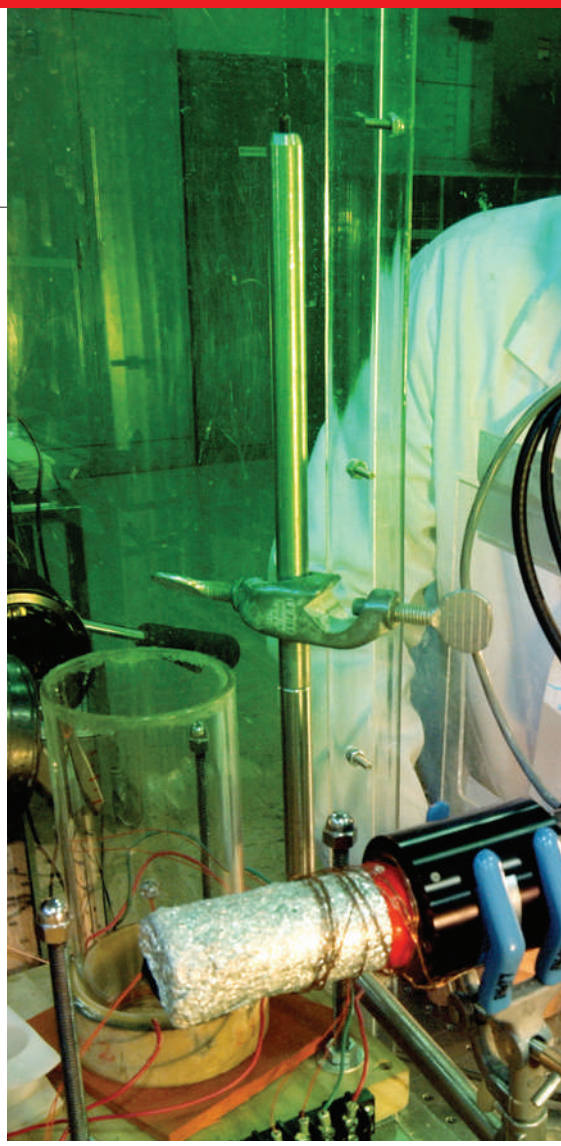
"They apparently narrowly focused the charge and avoided the question of whether the research was doctored," says Ken Suslick, a chemist at the University of Illinois at Urbana-Champaign who has been attempting to replicate Taleyarkhan's claims. Suslick is one of several researchers worried that Taleyarkhan's work may be fraudulent, and he wrote to Purdue about his concerns in June 2006. These include the apparent duplication of data between reports of supposedly independent experiments[4] (first raised by *Nature*), and a report[5] that the spectrum of neutrons that Taleyarkhan claims to have detected from bubble fusion exactly matches that of a standard radioactive source called californium. Taleyarkhan has since replied that when he measures neutrons emitted by californium in his lab, he finds something quite unlike what he sees from his fusion experiments[6]. But a recent preprint points out that Taleyarkhan omitted some of the original spectral data in his reply, and that the full data set still looks like californium[7].

The university never responded to Suslick's concerns. Peter Dunn, Purdue's associate vice-president for research, told *Nature* that he believes the university followed its procedures. He declined to comment on why he never replied to Suslick, or on whether evidence related to Suslick's concerns was forwarded to either inquiry. Purdue hasn't revealed the identities of the members of the second inquiry panel, but Dale Compton, a professor of industrial engineering at Purdue and a member of the first panel, says he has no recollection of being asked to consider the questions about Taleyarkhan's data.

"Purdue's finding is as mysterious as bubble fusion itself."

Lefteri Tsoukalas, who asked Purdue to investigate Taleyarkhan in February 2006, has called the announcement "an outrage". Tsoukalas was head of Purdue's nuclear-engineering school until he resigned in October 2006 in protest at the way the university was handling the concerns. He notes that the usual procedure for handling allegations of scientific misconduct is to hold a preliminary inquiry, then either proceed with an investigation or close the matter. That did not happen in this case; instead, the university ran a second preliminary inquiry. Apart from Tsoukalas, calls by *Nature* have failed to locate anyone who raised concerns about Taleyarkhan's work who was interviewed during either inquiry. "Purdue's finding is as mysterious as bubble fusion itself," says Tsoukalas.

Taleyarkhan, however, strongly defends the university's process. "Purdue University in my opinion and experience has conducted an extremely thorough review and with my full cooperation," he told *Nature*. "Allegations made in the press have been known to Purdue, and the administration's thoughtful conclusions have been well-articulated in their statements."

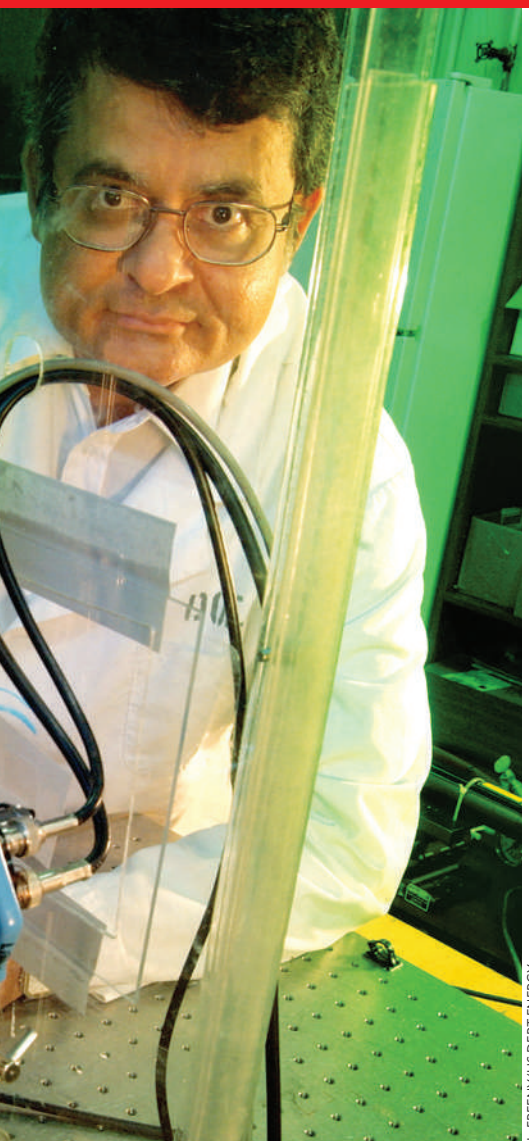


Rusi Taleyarkhan's controversial bubble-fusion findings have been upheld by his university.

Beyond the issue of misconduct, Purdue also makes clear that it endorses the scientific value of Taleyarkhan's work. In last week's press release, vice-president for university relations Joe Bennett stated: "Professor Taleyarkhan is engaged in very promising, significant research, and we hope he will now be able to give his full attention to this important work."

Seth Putterman of the University of California, Los Angeles, who has also been trying to replicate bubble fusion, thinks that Taleyarkhan's work is invalid. "Purdue's defence of Taleyarkhan's approach to scientific research taints their reputation," he says. "If Purdue were interested in maintaining their credibility they should have appointed external members to their panel." Mason, who is ultimately responsible for academic affairs at Purdue, did not respond to *Nature*'s requests for comment.

Purdue's announcement appeared on the



L. FREENY/US DEPT ENERGY

same day as Suslick, Putterman and others reported their attempt to replicate Taleyarkhan's claims in an experiment built to his specifications[8]. They did not find any evidence that fusion was occurring.

Suslick, Putterman and Taleyarkhan had received funding from the US Department of Defense's Advanced Research Projects Agency (DARPA) for a project to test Taleyarkhan's original claims. Taleyarkhan has spent around US$200,000 of this money on his bubble-fusion experiments. A spokeswoman for DARPA, Jan Walker, told *Nature* that although Purdue has not formally notified the agency of the inquiry or its results: "We are aware that an inquiry has taken place and are currently reviewing what, if any, action is required on our part." ■

Eugenie Samuel Reich

1. Xu, Y. & Butt, A. *Nucl. Eng. Des.* **235,** 1317–1324 (2005).
2. Taleyarkhan, R. P. *et al. Phys. Rev. Lett.* **96,** 034301 (2006) .
3. Reich, E. S. *Nature* doi:10.1038/news060306-1 (2006).
4. Reich, E. S. *Nature* **444,** 664–665 (2006).
5. Naranjo, B. *Phys. Rev. Lett.* **97,** 149403 (2006).
6. Taleyarkhan, R. P. *et al. Phys. Rev. Lett.* **97,** 149404 (2006).
7. Naranjo, B. Preprint at http://arXiv.org/physics/0702009 (2007).
8. Camara, C. G. *et al. Phys. Rev. Lett.* **98,** 064301 (2007).

BUBBLE FUSION
Find *Nature*'s March 2006 investigation into bubble fusion online.
www.nature.com/news/bubblefusion

Key biology databases go wiki

Barend Mons's first objective would be ambitious enough for most people: to meld some of the most important biomedical databases into a single information resource. But that's just the beginning. Mons, a bioinformatician at the Erasmus Medical Centre in Rotterdam, the Netherlands, also wants to apply the Wikipedia philosophy. He's inviting the whole research community to help update a vast store of interlinked data. If he and his colleagues can pull it off — and even the project's advocates are not sure they can — they could transform the databases that are central to the work of many life scientists.

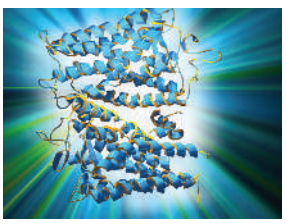
A test version of the project, provisionally dubbed Wiki for Professionals (www.wikiprofessional.info), is due to launch in the next month. It already contains data from key sources, such as protein information from Swiss-Prot and gene descriptions from Gene Ontology. Over the past year, Mons's team has woven together these and other archives to create what, from a user's point of view, seems to be a single database. The page on the muscular-dystrophy protein dystrophin, for example, contains data from Swiss-Prot together with links to disease information from the US National Library of Medicine, as well as explanatory text. Links to relevant publications in PubMed are also available.

Existing databases interlink to an extent, although the new resource is more comprehensive. But the next stage is the really radical bit. Biomedical research produces hundreds of thousands of papers a year, overwhelming database curators. To clear this bottleneck, Mons and his colleagues are allowing anyone to edit the entries, modifying and adding text and links as new work is published.

That's an attractive proposition, say database administrators. Michael Ashburner, a geneticist at the University of Cambridge, UK, helps run FlyBase, a collection of gene data on the model organism *Drosophila melanogaster*. The database receives around US$4 million a year from the US National Institutes of Health and employs up to five full-time curators, but still can't keep up with the relevant literature, says Ashburner, who is working with Mons on the new project. "We have a list of around 12 journals that we try to cover. Even that's tough."

"This will be a revolution."

Anyone motivated to register can curate Wiki for Professionals. Visitors to the dystrophin entry, for example, can update almost any of the information on the page, such as statements about the role of the protein in disease. Users can also start new pages, and from later this year will be given the option of creating pages for themselves, with links to relevant publications. A final function, and the one that most excites Mons, is the availability of text-mining software. This will allow users to probe links between proteins, genes and disease that may be revealed only by comparing a large number of papers and other data.



A. PASIEKA/SPL

Protein databases could be transformed by extra features.

"Mons is a visionary," says Amos Bairoch at the Swiss Institute of Bioinformatics in Geneva, a collaborator on the project and the creator of Swiss-Prot. "This will be a revolution."

Yet realizing the vision will be difficult. Top of the list of challenges is persuading the community to get involved. Adding one's own data is likely to be the biggest motivator — Bairoch and Ashburner say they get several calls a week asking for updates to databases, usually from researchers who want their own papers added. Whether this will be enough to keep the database fresh remains to be seen, given that employers and funders tend not to value updating information highly.

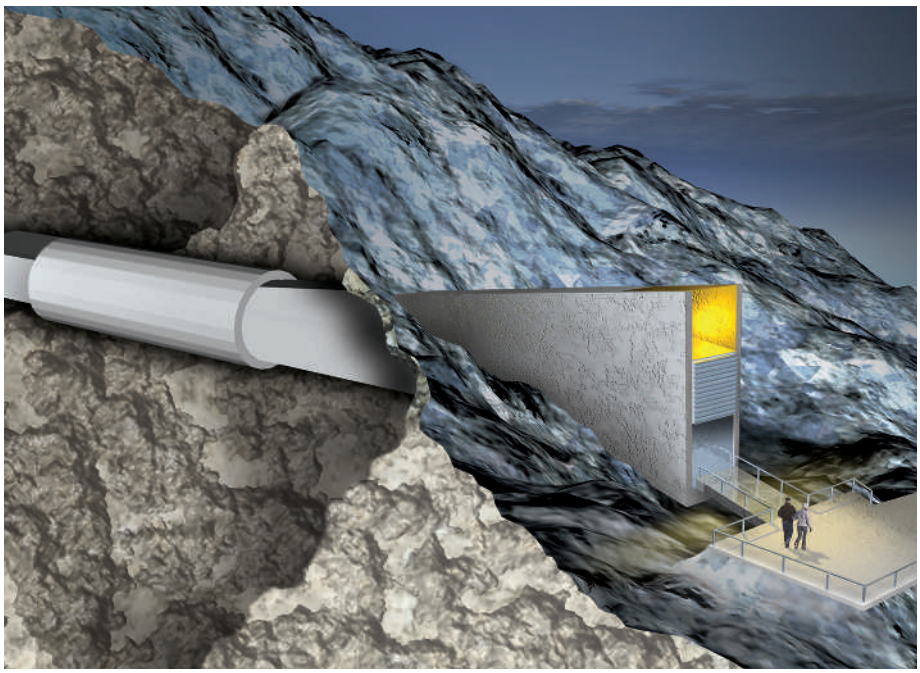
Wiki for Professionals will also have to ensure that additions don't just reflect individual researchers' pet theories. Mons hopes scientists will adopt entries relevant to their work and use automated systems to alert them to changes, which they can then amend if necessary. The original data in Swiss-Prot and other databases will also be protected.

The resource has been set up by Knewco, a scientific computing company based in Rockville, Maryland, and co-founded by Mons. The firm raised around $2 million in private funding to pay for the initial effort, and says basic access will be free. Revenue will be generated by charging drug firms and other users for premium services, such as the option to run a private version of the system incorporating proprietary data. ■

Jim Giles

Norway unveils design of 'doomsday' seed bank

GLOBAL CROP DIVERSITY TRUST



The Svalbard International Seed Vault will offer a safe haven for some 1.5 million crop strains.

The Arctic outcrop of Svalbard, just a few hundred miles from the North Pole, might seem like an odd place to install a freezer. But buried deep in the interior of an Arctic mountain, a US$5-million cold storage unit will house duplicate seeds for almost all of the world's estimated 1.5 million crop strains — protecting them from the ravages of such disasters as war and climate change.

The Norwegian government last week unveiled its design for the Svalbard International Seed Vault, nicknamed the 'doomsday vault' by the popular press. Construction should be completed in September, with the first seed deliveries arriving in March 2008.

Set at the end of a 120-metre tunnel cut into the freezing rock, where the natural temperature is –6 °C, the facility will be further cooled to –18 °C. "It will offer the best conditions for seed storage on Earth," says Cary Fowler, head of the Global Crop Diversity Trust, which will provide the facility's $125,000-a-year operating costs. The vault will even be insulated against climate change, says Fowler. Set 130 metres up the mountainside, it will be able to survive the most catastrophic projected rises in sea level.

Despite these precautions, calling the facility a 'doomsday vault' is disingenuous, argues Andreas Graner, who helps to run a seed bank at the Leibniz Institute of Plant Genetics and Crop Plant Research in Gatersleben, Germany. Even with optimal storage, most seeds lose their ability to germinate after about 20 years, so the facility is unlikely to bail out the human race from global disasters that might happen centuries in the future. To keep crop strains for long periods of time, seeds must be regularly grown into plants and a new generation of seeds produced.

"The vault will offer the best conditions for seed storage on Earth."

Although it will take more than a secure vault to save us from a global catastrophe, Fowler says it will act as a "global insurance policy", holding crop strains in the event that banks elsewhere are destroyed by local disasters. It will open for several weeks every year to take delivery of new strains from many of the 1,400 crop repositories around the world. In future, seeds lost in disasters like Typhoon Durian, which washed away samples in Jakarta, Indonesia, or in looting, as has occurred in Iraq and Afghanistan, can be replaced by dipping into the Svalbard vault. ■

Michael Hopkin

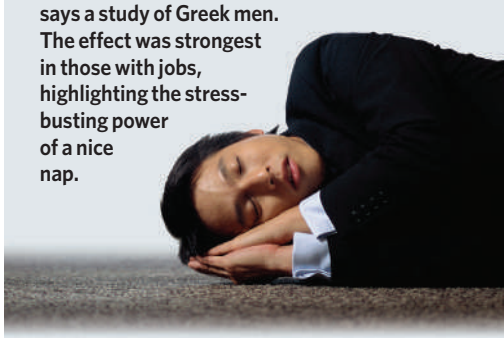
ON THE RECORD

"Sometimes you don't really trust men to pick out things you'll want to wear — I was very pleasantly surprised."

Harvard Medical School's Laura Mariani after her boyfriend bought her some molecule-shaped earrings (one serotonin, one dopamine) from jewellery company Made With Molecules. One to consider if you messed up on Valentine's Day.

SCORECARD

▲ **Siestas** Regular midday naps cut the risk of a death from heart disease by up to 64%, says a study of Greek men. The effect was strongest in those with jobs, highlighting the stress-busting power of a nice nap.



G. BROWN/CORBIS

▼ **Science books** Ten years ago, Britain's most popular non-fiction library book was Stephen Hawking's seminal *A Brief History of Time*. Now it's *You Are What You Eat*, a dieting tome by the widely criticized 'holistic nutritionist' Gillian McKeith.

NUMBER CRUNCH

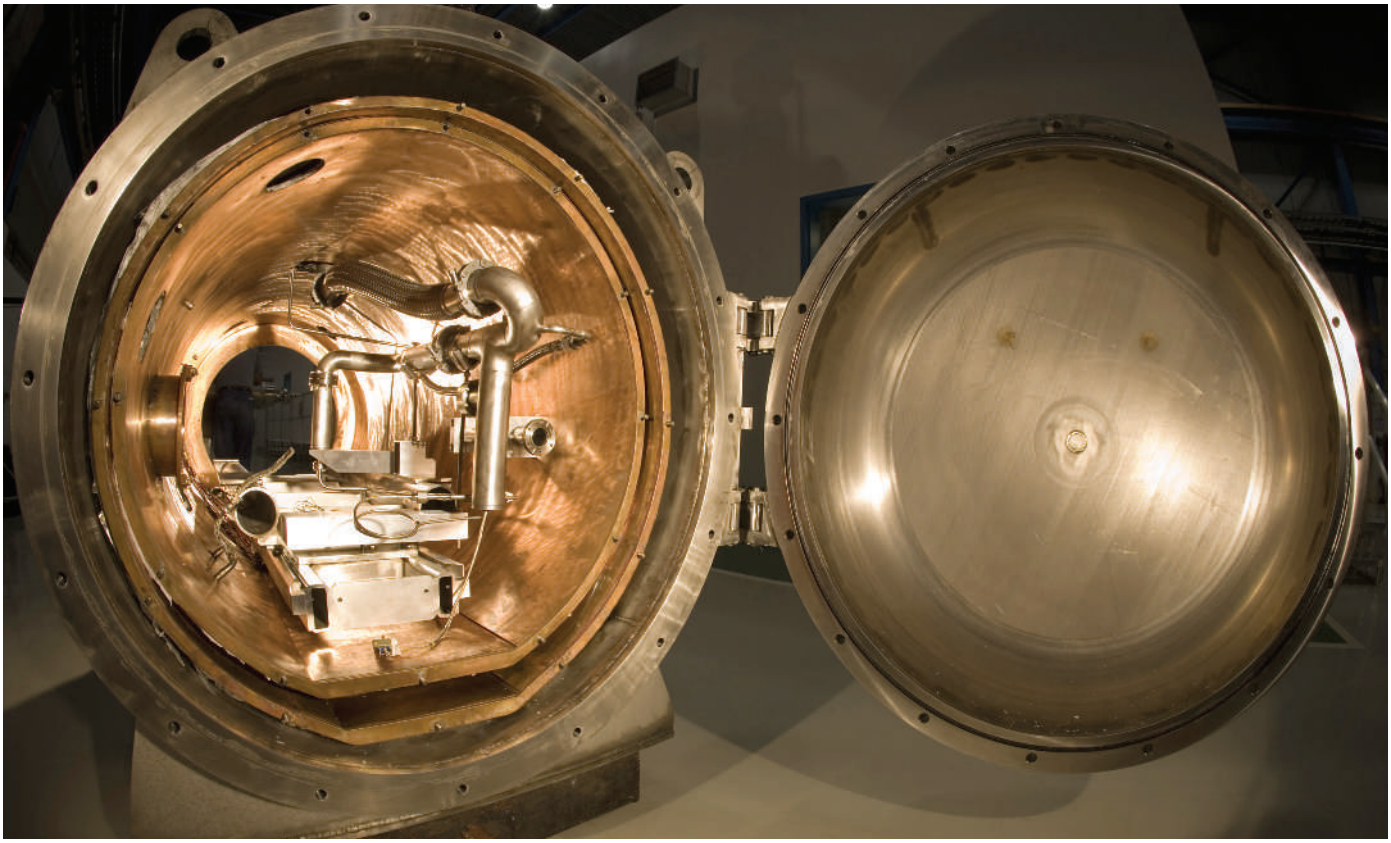
90% is the degree of certainty that global warming is being caused by human activity, according to the Intergovernmental Panel on Climate Change's fourth assessment report.

95% of Democrats who responded to a limited poll of both parties in Congress said they believed that it has been proved beyond reasonable doubt that human activities are warming the planet.

13% is the proportion of Republicans who responded who share that view.

Sources: Archives of Internal Medicine, BBC, IPCC, The Scientist

SIDELINES



FERMILAB

Starting from cold: many cryostat chambers will keep the superconducting International Linear Collider cool enough to accelerate subatomic particles.

Physicists pitch biggest accelerator

They have worked out the cost of a next-generation particle accelerator; now physicists must sell the dream machine to their governments.

At a meeting in Beijing, China, last week, a coalition of physicists set the construction price for the International Linear Collider (ILC) at about US$6.7 billion. That does not include the 22 million hours of labour the job would require, contingency costs or inflation. But the estimate is a first step in convincing governments to build the machine, says Barry Barish, a physicist at the California Institute of Technology in Pasadena and head of the design effort.

The ILC would be the most powerful particle accelerator in the world. Using superconducting cavities, the 31-kilometre machine would collide electrons and positrons at energies of up to 500 billion electronvolts. The resulting data might help to verify theories of physics beyond the standard model, which currently dominates the field. It would also complement data from the Large Hadron Collider (LHC) at CERN, the European particle-physics laboratory near Geneva in Switzerland, which is set to begin operation later this year.

But for the ILC to get any further, it must begin to gather political support from the United States, Europe and Japan. That could be tough. The United States has been wary of such projects since cancelling the Superconducting Supercollider in 1993, Europeans are busy finishing the LHC, and the Japanese government remains to be sold on the point of the machine, according to one official at the science ministry. "We can't provide financial assistance without being convinced about the purpose," he says.

"The $6.7-billion price tag does not include the 22 million hours of labour the job would require."

Even putting together a more detailed engineering design, the next step in the process, could be politically fraught. So far, the design work has been done voluntarily with little government money. The engineering phase will require governments to hand over substantial funding to international control, according to Burton Richter, a Nobel laureate and former director of the Stanford Linear Accelerator Center in California. And decisions will have to be made about what components and standards to use. "Somebody's got to have the authority to say 'It's this way, not that way,'" Richter says. "That gets into some interesting politics."

Where to build the ILC would be an even larger political battle. Europe, the United States and Japan have all shown varying degrees of interest in hosting the machine. In 2004, a similar dispute stalled ITER, a $6-billion experimental fusion reactor, for 18 months. "We hope not to end up in a situation like ITER," says Barish, although he points out that the ILC team has yet to solicit official proposals for where to site the accelerator.

The ILC's future hinges on what the LHC finds. If the LHC produces exciting results, political support for the ILC will be strong, Richter says. "If the LHC turns out nothing, I doubt very much that the governments are going to put up the money." ■

Geoff Brumfiel, with additional reporting by Ichiko Fuyuno

CAN PSYCHOLOGISTS HELP NASA?
Nature assesses how easy it is to pick personalities that survive high stress.
www.nature.com/news

Pigments help to date disputed masterpiece

Art collector George Lester Winward acquired this beautiful Madonna and child, now known as the *De Brécy Tondo*, in 1981 at a country-house sale in England. The more he studied it, the more he became convinced that it was painted by the Renaissance artist Raphael — not least because of its striking resemblance to Raphael's sixteenth-century *Sistine Madonna* in the Gemäldegalerie Alte Meister in Dresden. The art gallery, however, judged it to be a much later copy.

So Winward's foundation, the De Brécy Trust, sponsored a series of projects to generate evidence for his claim. It closed, not entirely conclusively, this week, with the publication of a laser Raman spectroscopic analysis of three tiny samples taken from the painting.

DEBRÉCY TRUST

Spectroscopic laser techniques suggest that the *De Brécy Tondo* is a Renaissance painting.

The analysis is consistent with the *De Brécy Tondo* being a Renaissance painting. For instance, one sample was shown to contain massicot, a popular yellow Renaissance pigment that was not used after 1700. By ruling out the presence of some alternative chemicals, the analysis also indirectly suggests a starch-based binding medium, commonly used in Renaissance paintings, and the use of the medieval dye turnsole for the blue robes. The presence of eighteenth-century Prussian blue pigment in discrete patches could be explained by recent touching up, says Howell Edwards, the University of Bradford chemist who did the analysis.

The trust hopes that these data, together with its convincing provenance research, will help persuade art historians, who rely to a large extent on visual assessment when authenticating paintings, that the work is indeed a Raphael, and thus worth millions. ■

Alison Abbott

www.debrecy.org.uk

Virgin offers big reward for capturing carbon

A US$25-million prize — one of the largest science prizes around — is on offer to the inventor of a device that will remove 'significant amounts' of carbon dioxide from the atmosphere.

"The winner must be able to demonstrate a commercially viable design which will result in the net removal of anthropogenic, atmospheric, greenhouse gases each year for at least ten years without countervailing harmful effects," state the rules of the Virgin Earth Challenge competition.

Among the judges are NASA climate scientist James Hansen; James Lovelock, inventor of the Gaia hypothesis; UK environmentalist and retired diplomat Crispin Tickell; and Australian conservationist and author Tim Flannery.

The competition is open for at least the next five years. British billionaire entrepreneur Richard Branson and former US vice-president Al Gore launched the prize on 9 February in London.

Historian to lead science forward at Harvard

Harvard University's new president, Drew Gilpin Faust, is a historian. But she is expected to continue many of the science-based initiatives that her predecessor, the controversial Larry Summers, put in place.

Plans are moving apace to create a science research complex in the Boston suburb of Allston — one of Summers' signature projects. He announced his resignation last February, in part over the furore about his comments on differences between men and women in science.

After her appointment on 11 February, Faust said that much remained to be done to address gender inequality in the sciences. She was the founding dean of the Radcliffe Institute for Advanced Study, a former women's college, which has put on science conferences on topics such as computational biology and tissue engineering. She also led two task forces looking at the status of women faculty at Harvard, including one on women in science and engineering, following Summers' controversial remarks.

Biologist Thomas Cech, head of the Howard Hughes Medical Institute in Chevy Chase, Maryland, was also a finalist for the president's job, but withdrew from the running last month.

K. SNIBBE/HARVARD

Pointing the way: Harvard boss Drew Gilpin Faust wants to address sexual inequality in science.

Horse genome to help human conditions

Twilight, the thoroughbred mare pictured here, is now the gold standard for all other horses: her DNA has been sequenced as the reference genome of the horse (*Equus caballus*).

The draft sequence, released on 7 February, should help the study of human conditions such as allergic disease, arthritis, exercise physiology and fertility. These studies will be aided by the fact that people have been breeding horses for 4,000 to 6,000 years and keeping close records of the bloodlines, allowing particular genes to be traced back through time.

D. ANTCZAK/CORNELL UNIV.

UK's Diamond synchrotron turns on the lights

Britain's Diamond synchrotron, a £260-million (US$505-million) device that is one of the country's largest pieces of new scientific infrastructure in decades, welcomed its first users late last month.

Diamond's initial users will study materials used in computer memories, a protein involved in cancer, and mineral samples from a meteorite that could shed light on conditions in the early Solar System. The facility currently has seven beamlines, producing X-rays with energies of between 100 and 20,000 electronvolts. It has funding to build another 15, which should all be online by 2011.

The debut of Diamond, situated near Oxford, marks the return of a world-class synchrotron to Britain: the first synchrotron was run in London in 1946. In 1993, government advisers warned that the country needed to build a new device to compete with systems elsewhere.

Catholic college sells land, but blocks stem-cell work

The University of Sydney in Australia will not conduct fetal stem-cell research in a new Aus$350-million (US$270 million) biomedical research centre planned on land obtained from a Roman Catholic college.

The 5 February compromise was reached after the university paid at least Aus$11 million to St John's College for land the Catholic school had controlled under a trust. St John's governing council sought the stem-cell restrictions even though the university had not planned such research at the new centre, which is expected to be completed by 2012.

University officials say they don't believe the agreement will create a precedent limiting stem-cell research at Australian public universities, such as Sydney.

Mars probe gets tunnel vision — in monochrome

The High-Resolution Imaging Science Experiment (HiRISE) on board the Mars Reconnaissance Orbiter — the newest and most powerful craft to arrive at the red planet — has lost its peripheral vision. And its colour vision is fading too.

Seven of HiRISE's 14 detectors are sending back spurious data, the mission team reports, and one of the four colour detectors has stopped working completely. This has led to only a 2% loss of signal so far, but the problem looks set to hit all of the detectors eventually.

"We do think it's a systematic problem for all of them," says Alfred McEwen, HiRISE's principal investigator, who is based at the University of Arizona in Tucson. "It's going to be a real irritant as it worsens."

Correction

The Editorial 'Light at the end of the tunnel' (*Nature* **445,** 567; 2007) should have referred to projects such as the Thames Barrier as 'adaptation' not 'mitigation'. In the related News story 'What we don't know about climate change' (*Nature* **445,** 580; 2007), the 2001 IPCC estimate for the range of sea-level rise should have read 9 to 88 centimetres, not millimetres.

A word in the right place

Lobbyists give an impetus to causes that can offer a significant advantage on Capitol Hill. But how do they go about getting their way? **David Goldston** examines their role in securing science funding.

On 9 February I had breakfast in Washington DC with four scientists who played a leading role in the recent report from the Intergovernmental Panel on Climate Change, to give them advice about testifying before Congress. The invitation came from the lobbying firm that represents the National Center for Atmospheric Research (NCAR), a privately run, largely government-funded institution in Colorado with which some of the researchers were affiliated. In the room, the number of lobbyists nearly matched the number of scientists. As well as a lobbyist from the private firm that represents NCAR, the institution's own government-relations person was present, as was someone from the government-affairs arm of the US Department of Commerce, for whom some of the scientists work and which provides some of NCAR's funding.

The scene captured the central role played by lobbyists in Washington — of particular relevance as the annual battle for the federal budget gets under way. Given the need to scramble for every available cent in a tight budget, the mind-boggling range and complexity of competing federal activities, and the byzantine nature of the appropriations process, lobbying is an unavoidable aspect of relying on funding from Washington.

Lobbyists concerned with funding for the physical sciences, for example, were in high gear last month while Congress was working out how much money to provide for the rest of the current fiscal year. For a while, it seemed as though funding for agencies such as the National Science Foundation and the Department of Energy's Office of Science would be frozen at last year's levels. But the resolution that emerged in late January from the House of Representatives, and that seems likely to be passed by the Senate, granted increases to those agencies (see *Nature* **445,** 572–573; 2007).

How did that happen? For a start, groups representing disciplinary societies and research universities began a letter-writing and meeting campaign directed at key players on spending bills including Speaker of the House Nancy Pelosi (Democrat, California) and Senate majority leader Harry Reid (Democrat, Nevada). A critical feature was

PARTY OF ONE

that the letters came not just from university presidents, who would be expected to seek funding, but also from business leaders who have not always been willing to make federal research funding a top priority. A few chief executives of leading companies may also have put in calls to Pelosi, which is the ultimate lobbying step and one rarely carried out when it comes to science funding.

The point of this activity was not to set out new arguments for spending more money on science. The case for funding basic research — that it leads to technological innovation and hence to economic prosperity — has changed little since 1945. Moreover, novelty of argument is rarely an advantage when pressing leaders to make quick decisions. The historian Carl Becker once remarked that Thomas Jefferson omitted original ideas from the Declaration of Independence because nothing would have been more foolhardy than trying to galvanize world opinion using ideas no one had ever heard of before.

Rather, the point of the lobbying was to make political leaders such as Pelosi feel that business leaders — people with political and economic clout outside Washington — cared about the funding. Given that the Democratic leadership already backed the idea of increased research support, as did the president, the additional pressure was enough to push congressional leaders into finding some additional money.

Targeting just a few officials was sufficient. Most congressional actions occur because a handful of well-positioned representatives or senators champion them, and no one else (at least of significance) feels strongly opposed. So getting Pelosi and a few key appropriators to push for higher funding was enough to carry the day.

That's not to say that a lobbyist for basic research can neglect the rank-and-file membership of Congress. Broad, even if shallow, support is needed to ensure continued funding. But there's no substitute for having a few ardent champions in the right places. That's how the ball got rolling on doubling funding for the National Institutes of Health in the late 1990s, with Senate supporters such as Arlen Specter (Republican, Pennsylvania) and Tom Harkin (Democrat, Iowa).

Like it or not, lobbying is now a necessity. Congress expects to hear from those who rely on government funding and is prone to interpret silence as lack of concern. Politicians can cut programmes — even those they view as beneficial — because they didn't hear from the proponents at a time when other interest groups were pressing their case.

That's one reason that interest groups waste no time in putting out press statements on the president's budget proposal as soon as it is released. Indeed, coalitions of industry and universities have been meeting for weeks to plot a strategy to push for funding increases for 'competitiveness' research in fiscal year 2008. The Washington representatives of schools and firms are supplementing their own work by hiring lobbying firms either for their strategic acumen or for their connections to key politicians. They envision a campaign through this budget cycle that is likely to involve not only calls and meetings 'inside the Beltway', but also advertisements and other ways to make competitiveness a compelling issue to voters nationwide.

Science advocacy groups are wise to try to avail themselves of every political tool. Competition for fiscal 2008 funding is likely to be intense even though congressional Democrats are generally more sympathetic than Republicans to domestic spending. Given the congressional Democrats' pledge to reduce the deficit, the overall budget 'pie' is not likely to be much larger than it was under the Republicans, but far more interest groups will be lobbying for a piece of it. ■

David Goldston is a visiting lecturer at Princeton University's Woodrow Wilson School of Public and International Affairs.

BUSINESS

France strives for sharper image

As US biotechnology companies seek to expand or relocate their operations abroad, Paris is pitching itself as a scientifically strategic location. **Rex Dalton** reports.

The many enticements of Paris are set for a difficult challenge: can they attract operating facilities from Californian biotechnology companies to move there?

Just three months ago, the French capital's regional government opened an office in San Francisco to urge entrepreneurial firms to choose to invest in Paris, rather than China, India or Britain. In a world where intellectual property can be moved instantaneously from place to place, laboratories can be built anywhere, and French officials would like to see more of them built in France. "We will be talking to everyone, from the life sciences to information technologies," says Frederic Le Roux, director of the new office.

The exercise is the latest example of strenuous efforts being made by France to build stronger links in science, technology and innovation with its veritable ally and occasional sparring partner — the United States. The moves attempt to overcome Americans' concerns about high costs, language and bureaucracy in France, by highlighting the excellence of its scientists and engineers.

Delegation skills

The international outsourcing of animal testing, clinical trials and even laboratory research has been a growing trend in Californian biotechnology firms. Indeed, academics from the state who have commercial ideas now often develop them through 'virtual companies', which have only a handful of US employees, and subcontract out their research and other operations, often to overseas contractors.

China, India and other locations in Asia are the main destinations for such 'offshoring'. And although Paris may be perceived in the United States as a cultural nirvana, French employment laws and high European salaries make it a less-than-obvious candidate for biotechnology expansion.

"France has an enormous advantage: it does very good research in great laboratories," observes Joseph Panetta, president of Biocom, San Diego's 470-member trade group. "But where I and my chief executive colleagues have questions, is in its ability to create and sustain an entrepreneurial environment like that of California."

French research chief Christian Bréchot: seeks closer ties with US scientists.

Collaboration between US scientists and their French colleagues is already widespread. And US biotechnology companies consider France's huge, state-run healthcare system to be an attractive environment for conducting clinical trials. Some companies have already set up shop there: PDL BioPharma of Fremont, California, for example, which is developing drugs for various illnesses including autoimmune disease and inflammation, opened up an office in Paris two years ago.

"There is a lack of willingness to take risks on venture capital and company formation."
— Joseph Panetta

But the Paris Regional International Mission Enterprise (PRIME) office in San Francisco is designed to take these interactions to a higher level. Its small team of staff will prowl university hallways and biotechnology firms in California and promote Paris's benefits over those of Bangalore or Shanghai. Tax breaks and legal assistance will be on offer to companies ready to make the move.

"Paris recognizes the unique productivity of the California biotechnology pipeline," says Matthew Gardner, president of the 350-member BayBio of South San Francisco. "Downstream, the possibilities are huge" for compounds and technologies that can be developed commercially.

The move by PRIME coincides with efforts by INSERM, France's main biomedical research agency, to strengthen scientific collaboration with the United States. An administrative structure has already been created for scientists from INSERM to keep their positions in France while they work abroad, and INSERM labs have been established in Kyoto, Japan; Haifa, Israel; Montreal, Canada; and Glasgow, Scotland.

In June, INSERM is expected to approve a proposal to station French scientists in labs at US universities. If approved, the first lab would employ 15–20 French researchers, specialize in neuroscience, and be directed by Emiliana Borrelli, an Italian neuroscientist who worked for 15 years in France and is now a faculty member at the University of California, Irvine.

Overseas experience

"I really believe that the future of INSERM research will rely on international expansion," says Christian Bréchot, director-general of INSERM. "The idea is to provide for the sustainable development of projects," he says, adding that participating INSERM scientists will be offered up to two four-year appointments, with a scientific review halfway through the eight years, he adds.

Borrelli's husband, Paolo Sassone-Corsi, is chair of the pharmacology department at the Irvine campus, and will be collaborating with the new lab. "Science should have no frontiers," says Borrelli. "But it can be hard for French scientists to move around. This new lab can help."

Two more INSERM labs, led by immunologist Jacques Banchereau at Baylor University in Dallas, Texas, and cell biologist Bruno Peault at the University of Pittsburgh, Pennsylvania, could also get the go-ahead in the months ahead.

But even as the scientific links grow stronger between biologists in France and the United States, US biotechnology leaders remain doubtful about France's attractions as a location for offshoring. "France has a very socialistic system," says Panetta, who once worked for a Californian firm, Mycogen, near Paris. "There is a lack of willingness to take risks on venture capital and company formation. They don't necessarily encourage entrepreneurialism; some discourage it." ■

A. DENANTES/ GAMMA, CAMERA PRESS LONDON

When two worlds collide

One half of a physics couple that met online, **Jennifer Ouellette** seeks some advice from married scientists on how to handle both long-distance and up-close relationships, while juggling career and family. Can love survive?

N. SPENCER

Last October, I became engaged to Caltech cosmologist Sean Carroll, capping six months of a long-distance romance that began via our respective physics blogs. Our his-and-hers blog announcements garnered the proverbial 15 minutes in the online scientific community, and it didn't take long before someone asked: "So, will you be relocating to California?"

Of course I will move to Los Angeles from Washington DC. Like any romantic, I would move mountains to be with my beloved; a cross-country trek, yowling cat in tow, is trivial in comparison. Sean is well worth that and more. But then I'm a self-employed science writer. You can give me a mobile phone, a laptop and a high-speed Internet connection, and I can do my job from almost anywhere.

Alas, scientists who marry scientists can't always get it together quite so easily. There is a daunting obstacle to be overcome: they must find jobs not just for themselves, but for their spouses. This is the 'two-body' problem: a reference to the challenge of calculating the paths of two objects interacting with each other. Mathematics solved the two-body problem long ago, but married scientists still struggle with it.

What little hard data are available show that they are in good company. According to several surveys of European scientists at least half of all scientists questioned have partners who are also working in science (H. L. Ackers *Gender, Mobility and Career Progression in the European Union: Final Project Report*; European Commission, Brussels, 2005). The problem is most acute in the natural sciences, says Londa Schiebinger, an expert on gender in science at Stanford University, who is heading up a US-wide survey of dual-career academic couples, building on a pilot programme at Stanford. Such a study is badly needed as there are very few hard US statistics on the matter — and those figures that do exist tend to be out of date. Schiebinger's group will survey more than 30,000 faculty members from the top US research universities, and conduct follow-up interviews and focus-group discussions.

One question the team hopes to answer is are such marriages tougher for female scientists? Besides being a minority in their field, female physicists struggle with the two-body problem more often than their male counterparts. A 1998 survey by the American Physical Society found that although only about 6% of its members are women, 43% of these are married to other physicists. In contrast, only 6% of married male physicists have a physicist spouse. Other studies have found that almost twice as many women chemists are married to or partnered with another chemist as

compared to their male colleagues, and 80% of women mathematicians are married to other scientists.

For most female scientists, the benefits of an intellectual connection with their partner probably outweigh any hardship. I certainly appreciate finding someone with whom I can discuss ideas, who continually challenges my assumptions and helps me view things from a different perspective; how much more true this must be for couples pursuing similar scientific careers.

According to Schiebinger, nirvana for married scientists in academia is two faculty (tenured or tenure-track) positions at the same institution or in the same area. Less desirable options include shared positions at the same institution, or one partner getting a tenure-track position while the other makes do with a lower-level lectureship or part-time position.

The course of true love

Back in 1976, physicist Ruth Howes didn't have that many options when she followed her husband Bob, a professor of dentistry, to Oklahoma. She took a temporary position, against the advice of her thesis adviser, and soon found herself unemployed. She worked part-time and focused on raising their children, but grew frustrated. "Nobody would hire me in Oklahoma," she says. The nadir came when a small private college refused to hire her because she insisted on teaching stellar evolution in her astronomy courses. "They didn't want any form of evolution taught," she says.

So when Ball State University in Muncie, Indiana, offered her a full-time position, she accepted, even though it meant living in different states while her children were quite young. "In those days, if you told people you were going to have a commuter marriage, they assumed you were getting a divorce," she says. "So we were a little ahead of the curve." But the Howeses made it work for 25 years by following two rules: "Talk every day, no matter what, and have a home for both partners on both ends. Both places should be home," says Howes, who now chairs the physics department of Marquette University in Milwaukee, Wisconsin.

Combining a family with a commuter marriage adds yet more complications, Howes acknowledges. As in physics, the many-body problem becomes much harder to solve.

For a year, Howes and her husband tried splitting their two children between the two households. From then on, the children lived with Howes while her husband did the commuting. She discovered that her children were very resilient. "They took it in their stride," says Howes, and became adept at packing. Although she worried about the potential psychological damage to her offspring, they didn't think it was so bad: "Every other weekend, we would basically freeze time. It was family time, and very special."

"Like any romantic, I would move mountains to be with my beloved."
— Jennifer

A few years ago, Bob Howes retired and joined his wife in Muncie. "That's something hardly anyone talks about: putting it all back together again," she says, admitting that initially there was conflict as the couple readjusted after so long apart. "The two-body problem is rough no matter how you look at it." They ended up buying a bigger house. Not only did it give each of them more space, it was "neutral ground". They also bought a second home in Santa Fe, New Mexico; remodelling that home in anticipation of Ruth Howes' retirement is a shared project.

Long-distance romance

Thirty years on, many scientific couples still opt for commuter marriages, at least at the beginning of their careers, rather than sacrifice one partner's dreams to the other's. A physicist friend of mine, Diandra Leslie-Pelecky, says: "If you both want to be high-powered researchers, you are limited in your choice of jobs, because there may not be many places with strong programmes in both areas." Now at the University of Nebraska in Lincoln, Leslie-Pelecky spent the first nine years of her marriage to fellow physicist Robert Hilborn commuting between Nebraska and the University of Massachusetts in Amherst. (Hilborn got a job at the University of Nebraska late last year.)

Do they have any advice for newlyweds, or soon-to-be-married scientific couples? "Both partners should win Nobel prizes," jokes Hilborn, thus giving them their pick of academic appointments.

For those of us whose last name isn't Curie, one or both partners must inevitably make concessions, and it might take longer than they would like to achieve their professional goals. Unlike Howes, Leslie-Pelecky opted not to have a family, a decision she is happy with. But she cautions that although both partners should be willing to make sacrifices, "if you compromise too much, you can limit your choices for future positions. The last thing you want is to have one partner feel that he or she got the raw end of the deal."

Chemist Julia Fulghum, of the University of New Mexico, Albuquerque, agrees. "We have tried hard to find positions that are a compromise for both of us, rather than ideal for one person and a bad fit for the other," she says. After spending a year apart from her husband and fellow chemist Stephen Cabaniss while he was a postdoc, they decided they didn't want a commuter marriage, especially as they knew they wanted children. Initially, their new positions were more limiting than they might have

J. EMBER

L. GERSTEIN

E. GERSTEIN

Edmund and Laura Gerstein on their wedding day (top inset) and during their manatee research.

liked, but over time, each established successful research and teaching programmes.

They now both have tenure. In fact, they've pulled this trick off twice. They found dual tenure-track positions at Kent State University in Ohio, before moving to their current jobs. But they had to make some trade-offs along the way. Both applied to a swathe of different academic departments, and they didn't put any geographical restrictions on their dual job search. That proved to be a key factor in their success, even though finding that first position took two years. In the interim, they both turned down attractive jobs at other schools, rather than live apart.

"Every couple has to figure out the issues that are most important to them," says Fulghum. "You have to be honest with each other about what is and isn't acceptable." So perhaps the Valentine cards have it right, love is communication. "We know couples that have made every possible combination work, and others who are miserable," says Fulgham.

> **"Every couple has to figure out the issues that are most important to them."**
> — Julia Fulghum

The flip side of the commuter marriage is the danger of too much togetherness, particularly for scientists who marry their bench partner. What if your careers mean you end up sharing office space or writing papers together?

The things we do for love

Edmund and Laura Gerstein have tested their togetherness to the extreme. They are married scientific collaborators at Florida Atlantic University in Boca Raton, specializing in animal acoustics. Not only do they work at the same institution, they once spent three-and-a-half years living in a small trailer behind a zoo — the better to study the acoustic behaviour of manatees. The couple endured extreme close quarters, no private bathroom, bizarre hours, and the occasional rampaging elephant knocking into their trailer. "It got to the point where we didn't really have to talk, we could just kind of grunt at each other to communicate," says Edmund.

They sometimes joke that it's a miracle they are still together, but during the experience they figured out an efficient division of labour that Edmund says helped their research and relieved the inevitable tensions produced by constant togetherness. For instance, Laura handled the computational aspects of the project while Edmund worked with the animals, putting some much-needed distance between them, at least during work time.

Does having similar research areas help or hurt a job hunt? Fulghum believes too-similar fields can be more of a disadvantage for younger, less established scientists. "I've frequently observed a sometimes conscious, sometimes subconscious, tendency for faculty evaluating two junior people to assume that only one of them can be 'good,' or that they have to figure out which one is 'best,'" she says, adding that this is less of a problem at the senior level because you are judged more on accomplishments.

For Fulghum and Cabaniss, the overlap proved advantageous, because Kent State was interviewing for two positions, and their research areas were sufficiently different: she works on materials characterization, and he specializes in environmental geochemistry. The University of New Mexico's policy encourages hiring spouses if one member of the couple is being actively recruited, and the relevant departmental heads worked to bring about the dual positions. Yet the couple also interviewed at less progressive universities that "made it very clear they were not interested in having a couple in the department", says Fulghum.

One rarely discussed aspect of the two-body problem is divorce rates. Certainly there have been scientific couples, some quite prominent, for whom the challenges proved too great. But the frequency of such breakdowns is unknown.

If all this anecdotal evidence proves anything, it's the need for the comprehensive Stanford survey currently under way. A report on the findings will be released later this year. The hope is that it will provide data to back up personal experiences, so that universities can formulate the best policies for their married faculty members.

In the meantime, I've gleaned some useful titbits of advice and encouragement (see 'Tips for newlyweds') for my own foray into marriage. No doubt there will be a few bumps in the road ahead as Sean and I adjust to life together, but we're ready to take the next step. Fortunately, we're both good communicators, as our blogging activities and six-month bicoastal love affair show. And we won't have the two-body problem. That gives me confidence in our shared future; the rest — well, it's mostly logistics. ■

Jennifer Ouellette is a freelance writer currently based in Washington DC.

Practical view: family life for the Fulghums came before their careers.

Tips for newlyweds

If you're looking for a dual appointment

- ♥ Be willing to make some compromises. Make sure you agree on what's acceptable and what's not in your careers and your family life.
- ♥ Publish. Then publish again. The more brilliant the candidates, the easier it is to place them.
- ♥ Be active in professional societies to gain recognition in the wider research community.

If you're going to have a 'commuter marriage'

- ♥ Communicate. It's important to talk every day, no matter what.
- ♥ Both partners should feel at home in both cities, with belongings in both locations.
- ♥ Make your time together count by clearing your respective schedules.

If you want to start a family

- ♥ Alternate your work or teaching schedules so that one partner is always available to stay home with a sick child. This saves on childcare costs.
- ♥ Set aside 'family time' so the children build healthy relationships with both parents.
- ♥ Look for a department with 'family friendly' policies and a supportive infrastructure.

If you're putting it back together after years of commuting

- ♥ Make sure each of you has a private space in the home where you can retreat if necessary.
- ♥ Expect some friction at first, as you adjust to the compromises of communal living.
- ♥ Consider buying a new house, or embarking on a joint project in which you are building your future together. **J.O.**

Share your valentine stories on *Nature*'s newsblog at http://blogs.nature.com/news/blog/2007/02/scientists_in_love.html

KILLER IN THE KELP

Could a change in the dining habits of orcas crash an ecosystem? **Mark Schrope** reports on a mystery that reveals how little we know of the oceans.

PACIFIC STOCK/PHOTOLIBRARY.COM

When you're up to nine metres in length, weigh more than an African elephant and can swim at over 50 kilometres an hour, you can expect to be admired for your sheer brawn. Striking black-and-white markings over a sleek, streamlined torso will earn you points for beauty. But orcas, also called killer whales (*Orcinus orca*), can lay claim to brains too. These magnificent creatures have devised cunning methods to earn a top-carnivore's living from the sea. Some force sharks to the surface and club them with their flukes; others hunt the sharks down in underwater gangs. In Norwegian fjords, the orcas herd up herring, while on the shores of Patagonia they all but beach themselves to pick off seals. These differences in hunting practices between the orcas, along with differences in their looks and songs, have led some experts to suspect that the creatures actually belong to several separate species (see 'Species apart').

An abiding mystery, though, is whether a change from one feeding habit to another could profoundly alter the balance of marine ecosystems. In 1998, a team led by marine ecologist Jim Estes at the University of California, Santa Cruz, proposed that just such a shift might explain an enigmatic and precipitous decline in Western Alaska's population of sea otters (*Enhydra lutris*)[1]. The decline was of particular interest because the lack of predatory sea otters caused a boom in the sea-urchin population; the flourishing sea urchins, in turn, laid waste to large areas of kelp forest, thus changing the balance of a whole ecosystem.

Five years later, Alan Springer, a marine ecologist at the University of Alaska in Fairbanks, and colleagues, including Estes, took the hypothesis several steps further. They suggested that the sea otters' demise might be the last stage of a grander collapse in which orcas had shifted repeatedly to new prey as old prey ran low[2]. The cause of this change, they argue, was commercial whaling, which deprived some orca populations of the great whales — such as humpbacks (*Megaptera novaeangliae*) and blues (*Balaenoptera musculus*) — on which they used to feed. The orcas turned instead to smaller sea mammals such as harbour seals (*Phoca vitulina*) and sea lions (*Eumetopias jubatus*), eventually working their way down to the otters and triggering the destruction of the kelp forests. The idea that whaling could have changed the orca's diets had been suggested before by French researchers[3], but this was the first time that it had been linked to wholesale ecological change.

> **"We see literally tens if not hundreds of humpbacks, and we just never see orca attacks."**
> — Paul Wade

"I'll be the first to admit that it isn't even close to being definitive," says Estes, who came to the story through his studies of sea otters "It was intended to be provocative." Mission accomplished, judging by the two rebuttals contained in an upcoming issue of *Marine Mammal Science*[4,5]. "Superficially attractive" is pretty much the nicest term used — "simplistic and highly selective" and "poorly supported" also feature. "It's a beautiful idea. I wish it were true," says Lance Barrett-Lennard, a biologist studying orcas at the Vancouver Aquarium in British Columbia, Canada. "But it's wrong." Meanwhile, Springer and his colleagues are working on their own rebuttals.

The detractors challenge nearly every link in the chain of evidence that implicates the orcas, starting with whether they ever actually ate many great whales. There is little doubt that they ate whales: Springer, Estes and their colleagues point to historical records from whalers recount-

Species apart

Rats and humans are said to be the planet's two most widespread mammals, but orcas run a fairly close third. Found in every ocean and in every clime, orcas have a wide range of tastes in prey. Yet ever since Carl Linnaeus gave them the name *Orcinus orca* in 1758, they have been treated as a single species. Now, opinions are starting to change. "Everybody acknowledges we will soon have a revision of [orca] taxonomy," says Lance Barrett-Lennard, a biologist studying the mammals at the Vancouver Aquarium in British Columbia, Canada. "But if we do it now, we'll make a mess."

One of the potential re-classifiers is Robert Pitman, a biologist with the Southwest Fisheries Science Center in La Jolla, California. After countless hours in helicopters, small boats and ships scouring the frigid waters around Antarctica, he may be close to amassing evidence that is strong enough to lead his peers to recognize at least one new species of orca.

The discussions of a potential new species of orca go back to at least the 1980s, when two separate teams of Russian scientists attempted to describe a new species in Antarctic waters[8,9]. The work was generally discounted because of insufficient data, and in one case because samples were apparently lost in a museum flood in Vladivostok. But Pitman, who was already involved in Antarctic surveys for the International Whaling Commission, thought there might be something to the Russian claim. "I started looking for differences," he says, "and found evidence fairly quickly that there were at least three different recognizable types down there." Ingrid Visser, founder of the Orca Research Trust in Whangarei, New Zealand, has also observed distinct populations around Antarctica, as have other researchers.

R. PITMAN

Sharp shooting: blubber samples taken from the orcas' sides with a crossbow could provide proof of a species divide.

Not so black and white

One form, type A, has the classic black-and-white look, is typically found in open water, feeds mainly on minke whales (*Balaenoptera acutorostrata*), and has males that are 7–8 metres long. The other two forms have distinctive songs, are smaller than type A, and are found near shore in pack ice. Type B feeds on seals, and possibly also on larger whales, whereas type C seems to feed exclusively on the Antarctic toothfish (*Dissostichus*

ing orca attacks. Television viewers around the world have seen the harrowing sequence of a grey-whale calf being eaten by orcas that featured in the documentary series *Blue Planet* — not an uncommon occurrence. In fact, the very name 'killer whale' may be a confused translation of the old Spanish term *asesina ballenas*, or whale killer.

Biting attacks

The problem is that in some of the places where whale populations have rebounded since commercial whaling stopped, biologists have yet to see any orca attacks on whales. "We can be out there on the water day after day with literally tens if not hundreds of humpbacks, and we just never see attacks," says Paul Wade of the National Marine Mammal Laboratory in Seattle, Washington. But whales are sometimes found with scars from orca bites. "If they aren't eating them, why are they biting into them?" asks Estes. Although no current population of orcas has been seen hassling humpbacks, he says, that doesn't mean that they aren't doing so — or did not in the past.

There is also a possibility, raised by Hal Whitehead of Dalhousie University in Nova Scotia, Canada, that whaling helped the orcas, rather than cheating them of their food. Harpooned whales that were left floating on the surface were often partially eaten by orcas, he and his colleague Randall Reeves note in their paper on the subject[6]. In the heyday of whaling, harpoon shots may have sounded like dinner bells to orcas' ears, announcing a big fresh meal that required no hunting. In this version of the hypothesis, the end of whaling changed the orca's diet not because the number of humpbacks hit an all-time low, but because whalers stopped providing orcas with ready-to-eat meals.

"The debate basically just highlights how difficult it is to study these animals."
— Terrie Williams

The next step in the cascade is no less controversial. Wade points out that although populations of harbour seals crashed around the Aleutian Islands off Alaska in the 1970s and 1980s, they were stable in parts of the Bering Sea where commercial whaling had been just as heavy. And questions remain as to whether the seals and sea lions declined species by species or all at once.

At least as far as the North Pacific and Bering Sea are concerned, Estes agrees that the issue of sequentiality is important. Sequential collapse would point very specifically to the orcas, whereas simultaneous collapse might reflect human exploitation of the fish that all the mammals prey on. But simultaneous collapse does not rule out variants of the hypothesis. At the other end of the world, Terrie Williams, also of the University of California, Santa Cruz, and her colleagues have proposed that the simultaneous collapse of the southern elephant seal (*Mirounga leonina*) and southern sea lion (*Otaria flavescens*) populations in the Southern Ocean might have been caused by orcas finding new prey after the end of whaling[7].

The 'fewer sea otters more sea urchins' link in the chain is not quite as controversial. Orcas have been seen to eat sea otters, in one case sweeping a group off the ice it was lying on to catch the otters in open water. But the evidence that enough of this sort of thing goes on is only circumstantial. Proponents point to observations that populations of sea otters have declined in areas of open water that contain orcas, but not in nearby orca-free lagoons, for instance. "There's nothing wrong with circumstantial evidence," says Wade. "You just have to be clear that it is circumstantial evidence, not direct evidence."

A, B, C: illustrations of the three orca populations.

U. GORTER

mawsoni). Both type B and type C are more grey and white than black and white. The habitats of all three orcas overlap to some extent during the summer — although their movements across the year are largely unknown — but no evidence has been seen of interbreeding between the three populations.

In a paper to appear in the *Journal of Mammalogy*, Pitman and his colleagues describe a helicopter survey of 221 type-C orcas — the population they think is the most likely to be recognized as a separate species given the information gathered so far. For more evidence, though, the team is finishing genetic analyses of more than 50 tiny blubber samples that have been pulled from orcas' sides with a crossbow. If the genetics prove their point, then type-C orcas will become the first new whale species named since 2003 (ref. 10).

This would not be a purely academic point. Commercial fishing for Antarctic toothfish is expanding, and if it proves to be the only known prey for a new species of orca, then that would have implications for conservation. Although orcas are treated as 'conservation dependent', they are not considered endangered. But a relatively small population newly classified as a species might be.

You are what you eat

The need for genetic data is universally agreed, as neither behaviour nor size is a sure guide. Take, for example, the question of fish eaters versus mammal eaters. In the North Pacific Ocean and the southern Bering Sea, two groups of orcas live near the shore, one of which eats mammals, the other mainly fish. The same distinction is seen in the Antarctic. For a while, says Barrett-Lennard, a similar distinction at both ends of the world made it seem possible, even likely, that the behaviours were those of two different species. But the genetic tests said no. "We went, 'Holy smokes! Our mammal-eaters here are not genetically related to mammal-eaters elsewhere'," says Barrett-Lennard. That pretty much ended any hopes that behavioural differences would be enough to define and separate orca species. "The search right now is for consistent, deeper divisions," Barrett-Lennard stresses, pointing to the need for data from less-studied areas such as the central Indian Ocean and the Japan Sea.

Pitman, for his part, says that his group will make conclusions based only on what the data tell them. Still, he argues that Antarctic orcas are a special case, in part because populations there are the only ones known that can be easily separated even by untrained observers. "I think Antarctica is particularly interesting because the morphology of the whales is so divergent," he says. Visser agrees. "It is just so graphically, graphically clear," she says. "I knew when I saw them the first time."

Results of the genetic tests are due any day now. And whatever those tests find, they should help to move the debate onward. "Come on," says Visser, "it's time that we take a step forward." **M.S.**

Look out for orcas: Western Alaska's sea otters may be a new and vulnerable prey.

D. BOAG/PHOTOLIBRARY.COM

That said, you don't need many orcas eating otters for the effects to be felt. As big, fast, warm-blooded creatures, orcas need a phenomenal amount of energy. Williams calculates that relying on sea otters for even a fraction of that energy would mean eating a great many of them. But orcas eating otters would not prove that the cascade effect suggested by Springer and Estes exists. Wade thinks that a turn to sea otters as food would not require a collapse in the stocks of other possible prey. Hunting practices are learned, so if, for example, an injured mother decided to eat sea otters because they were easier to catch, that behaviour could spread in a population irrespective of what other food was available. This might also reconcile the possibility that orcas used to feed on humpbacks with the claim that they no longer do.

Williams, like some other participants, has been surprised at how long the debate has gone on. "It basically just highlights how difficult it is to study these animals," she says. Clearly, one of the reasons that the arguments have been so heated is that the conservation stakes are high. The domino effect from whales through seals to sea otters as a potential explanation for the declines in marine mammals could take some of the pressure off commercial fishing. It might also be seen as a strong argument to continue the whaling moratorium — although some fisherman have argued that the orcas should be culled. The debate is likely to continue for years, meaning that orcas will remain a symbol of how much remains unknown about the oceans. "You have to have the debate and you have to have people thinking creatively, and then, ultimately, you'll come down to the truth," says Williams. "It will happen — we just don't have the information now to know what the truth is." ■

Mark Schrope is a science writer in Florida.

1. Estes, J. A. *et al. Science* **282,** 473–476 (1998).
2. Springer, A. M. *et al. Proc. Natl Acad. Sci. USA* **100,** 12223–12228 (2003).
3. Barrat, A. & Mougin, J. L. *Mammalia* **42,** 143–174 (1978).
4. Trites, A. W. *et al. Marine Mammal. Sci.* advance online publication, doi:10.1111/j.1748-7692.2006.00076.x (12 July 2006)
5. Wade, P. R. *et al. Marine Mammal. Sci.* advance online publication, doi:10.1111/j.1748-7692.2006.00093.x (21 December 2006)
6. Whitehead, H. & Reeves, R. *Biol. Lett.* **1,** 415–418 (2005).
7. Branch, T. & Williams, T. in *Whales, Whaling and Ocean Ecosystems* (eds Estes, J. A. *et al.)* Ch. 20, 260–278 (University of Cambridge, California, 2007).
8. Mikhalev, Y. *et al. Report of the International Whaling Commission* **31,** 551–556 (1981).
9. International Whaling Commision *Report of the International Whaling Commission* **32,** 617–631 (1982).
10. Wada, S. *et al. Nature* **426,** 278–281 (2003).

A breath of fresh air

How often does independent research change laws as well as minds? A lobby group in Delhi is forcing the Indian government into new regulations. **Apoorva Mandavilli** meets its leader.

M. SWARUP/AP

A decade ago the city of Delhi was choking. Fumes from the growing traffic rendered the air thick and foul with toxic chemicals, earning India's capital city the dubious distinction of being the fourth most polluted city in the world. Levels of fine particles in the air were nearly 17 times higher than the permissible maximum. You could almost feel them as you breathed.

Visit Delhi today, and the difference is palpable. Green-striped buses and auto rickshaws rush past powered by compressed natural gas. Levels of sulphur in diesel have been brought down from 2,500 parts per million to 500 parts per million. Concentrations of particles in the air are still three times the national standard, but more bearable — the air feels unmistakably cleaner.

The improvement is largely due to the efforts of one small non-governmental organization, the Centre for Science and Environment (CSE). Founded by the science journalist Anil Agarwal in 1980, the Delhi-based group launched a relentless campaign in 1996 to replace diesel in Delhi's public transport with a cleaner fuel: compressed natural gas. Its headline-grabbing tactics were what you might expect from a group founded by a science journalist: at one point it hired a booth at a Delhi car show and offered attendees lung tests. In April 2002, after years of legal battles, India's Supreme Court forced Delhi's public vehicles to switch to compressed natural gas. "It's undoubtedly one of the most influential organizations in the country," says Mahesh Rangarajan, a former Rhodes scholar and commentator on Indian politics based in Delhi.

So how did a small band of campaigning journalists evolve into a respected environmental pressure-group powerful enough to change laws and send multinational companies running for cover? The CSE is that rare entity, an activist group that is prepared to back up its campaigns with its own research. India has countless problems, but the CSE picks its fights wisely. For the past 20 years, it has focused on five main areas: air pollution, climate change, water management, pesticides and poverty eradication.

> **"We have the arrogance to believe we are as powerful as government."**
> — Sunita Narain

Today the CSE has a charismatic woman at its head in Sunita Narain, who took over the reins of the organization in 2002 after the death of Agarwal. "We are essentially playing the role of a watchdog, pushing for policy and being a public advocate," she says. "We have the arrogance to believe we are as powerful as government."

Most recently, the CSE has taken on the soft-drinks giants Coca-Cola and PepsiCo. In August 2006, the CSE released a technical analysis of 12 popular soft drinks made by these companies and sold in India, claiming that they contained toxic pesticides, including lindane, DDT, malathion and chlorpyrifos, at up to 36 times the European standards for bottled water.

In a testament to the organization's credibility, four Indian states promptly banned sales of soft drinks made by Coca-Cola and PepsiCo at schools, government-owned offices and hospitals, with one state imposing a total ban on sales and production. Sales of Coca-Cola products nationwide fell by about 18%.

As in other CSE campaigns, the battle was waged on two fronts. Articles reporting the results appeared in the CSE's fortnightly magazine *Down to Earth*, while the technical data generated by three staff scientists were posted on the CSE's website.

"To my mind, that's what makes them powerful," says David Dickson, director of UK-based SciDev.net, an online science information resource for the developing world. "I'm not saying all their science is perfect, but I think they realize the importance of having science behind

their argument. That differentiates them from many other environmental groups."

Agarwal played the research card from the start. A mechanical engineer by training, Agarwal's career as a journalist stretched from Indian newspapers to UK publications including *Nature* and *New Scientist*. Returning to India in 1980, Agarwal launched the CSE with two fellow journalists. They began by researching environmental issues, looking at poverty and resource management at the village level and putting together the first of the ambitious series of 'Citizen's Reports' published under the general title *State of India's Environment*.

At that time, environmental issues were discussed only in developed countries, or as Narain puts it: "First you get rich, then you pollute, then you get dirty, then you start cleaning up." But Agarwal argued that India was too poor not to care about the environment. "He was putting environmental issues on the development agenda and he was doing it at a time when most people weren't," says Dickson.

Hard graft

The material for those first reports was gathered painstakingly. Narain, who joined the CSE in 1981 as a volunteer, and her colleagues collected newspaper articles, literature from non-governmental organizations and scientific papers. They also wrote to scientists for more information. "The task was to take all that, assimilate it, analyse it and find a trend. It was massive," recalls Narain.

Right from the first report in 1982 it was apparent that the CSE had found a winning formula. The report gained accolades from the international press, stirred Tunisia to launch its first environmental ministry and inspired Lester Brown of the Worldwatch Institute in Washington DC to launch his annual *State of the World* reports.

A 1991 report by the CSE linking monsoon flooding with deforestation also helped transform the environmental debate in India. Although the CSE wasn't the first to make the link, its report consolidated research and traditional knowledge and presented it in accessible language, says P. V. Unnikrishnan, humanitarian coordinator for ActionAid in Asia. "In 1991 for such a report to come out, touching on such a controversial issue, was very good. It helped kick-start an informed debate."

Although generating its own research gets the CSE attention, it also means that its research can become the focus of criticism. For example, its reports on pesticides in soft drinks were attacked on technical grounds from the start by both the multinationals involved and the Indian government. This high-profile fight has earned Narain the moniker of 'Coke lady' and has gained the CSE fame beyond Delhi.

Standard bearer

Narain is used to public criticism, but she admits the soft-drinks fight has been much nastier than she expected — and the CSE got into it almost by accident. The organization became interested in general water-safety standards in 2002, when it tested ten brands of bottled water and found pesticides in them. Puzzled by the vehemence with which Coca-Cola and PepsiCo, which had some of the purest bottled water, were fighting demands for water standards, the CSE decided to look at pesticide levels in soft drinks. "We just tested out of curiosity, but now that we're in it, we can't back off," Narain says.

Nevertheless, Narain says her main gripe isn't with Coca-Cola or PepsiCo, but with the government, which has dragged its feet in setting safety limits for pesticide residues in soft drinks. "Even now, our only demand from the government is standards," says Narain. "We found something that's unclean. Clean it up. We want standards. Simple."

Most countries, including India, have standards for pesticide levels in the water that goes into carbonated drinks but not for the product itself. That may be acceptable in countries where the raw material is relatively clean, but not in India, Narain argues.The solution, she says, is for India to implement testing of the final products.

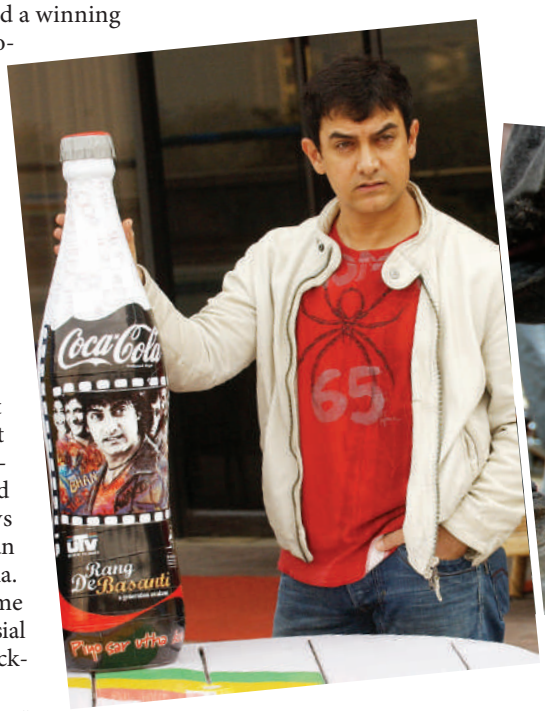
The CSE released its initial findings on pesticides in soft drinks in 2003. After initially disparaging the group's methods, the government set up a Joint Parliamentary Committee — only the fourth in India's history — to investigate.

The committee upheld the CSE's methods and findings and asked the Bureau of Indian Standards to review soft-drinks standards. Two years later, the bureau set new pesticide limits, which the CSE welcomes. But it has yet to make them law — thanks, Narain alleges, to interference from the soft-drinks industry.

Imperfect harmony

Coca-Cola and PepsiCo are both standing by their products, saying that their soft drinks made in India are of the same quality as anywhere else. The companies maintain that setting standards is complicated because there are no reliable tests to assess pesticide levels in carbonated beverages. "Measuring for the pesticides [in a final product] is very difficult to do," says Kari Bjorhus, a spokeswoman for Coca-Cola in Atlanta, Georgia.

The CSE data have not been peer reviewed but scientists contacted by *Nature* say that the organization followed methods accepted by bodies such as the US Environmental Protection Agency. "The data are believable and the levels they find [of pesticides] are believable," says Laura McConnell, a research chemist for the US Department of Agriculture's research service. "These should be



M. ROMANA/AFP/GETTY; A. SOLANKI/AP

Coca-Cola has recruited actor Aamir Khan (left) to boost its image, while students protest by smashing bottles.

values that would stand up in court as valid." McConnell also questions the soft-drinks companies' claim that testing for pesticides is complicated. "I don't think it's that complex," she says.

After the 2003 report, Coca-Cola approached the UK government's Central Science Laboratory (CSL), which tested 180 samples and reported that pesticide levels in the soft drinks were below "acceptable" limits for drinking water. But, Narain says, the lab never released the detailed reports or chromatograms for scrutiny. Bjorhus confirms that the CSL did not release its full data to the public, but she argues that the lab didn't find any pesticides, "so there is nothing to review in terms of findings". She notes that the CSL is "one of the most respected laboratories in the world and their integrity is unquestioned".

The UK laboratory's report has not yet persuaded Indian state governments that the drinks are safe, and they remain banned in many schools and state-run institutions. The pesticide battle has also fuelled a wider resentment in India against multinational companies, with people smashing up drinks bottles. In response, Coca-Cola has launched an expensive media campaign, featuring top movie stars such as Aamir Khan vouching for the drinks' safety. But Narain says, "their sales are still down, Aamir Khan or no Aamir Khan".

Close focus

Away from the media battle, the technical arguments over safety are now focused on the validity of different tests. After the CSE released its 2006 report, assessing 25 new soft-drink samples for pesticides, PepsiCo's president Indra Nooyi visited India in December, saying that the two multinationals would work together to find "a breakthrough science-based method to reliably and consistently measure the levels in finished products".

> **"We are not asking for closure of Coke and Pepsi. We just want standards."**
> — Sunita Narain

Coca-Cola and PepsiCo have commissioned AOAC International, an independent not-for-profit American scientific association, to review about 17 available analytical methods and to make recommendations to the Indian government. The report is expected in February.

Narain maintains that she would be willing to accept evidence showing the CSE's science to be wrong, but that the companies or the government have to provide it first. "We are not asking for closure of Coke and Pepsi, we are not interested in them as companies," says Narain. "We just want standards."

Today the CSE employs 125 journalists and researchers, about half of whom hold degrees in science. But it wasn't until 1992 that Agarwal launched the magazine *Down to Earth*, backed by the belief that journalism is one of the most powerful ways to change things in India. Today, the magazine has a popular website and 20,000 paying subscribers. Like Agarwal, Narain sees herself as a journalist first, and is not afraid to attack those scientists who, she says, stay on the sidelines or sell their expertise to companies. She once wrote an editorial criticizing the scientific community for its "arrogance and ignorance, both completely unacceptable".

Scientists have in turn criticized the CSE. In 2005, Narain was asked to head the Tiger Task Force, which looked into the controversies inherent in human and tiger coexistence and advised on tiger conservation. The final report left a lot to be desired, says biologist Ullas Karanth of the Wildlife Conservation Society, India, but he was impressed with Narain's hard work and willingness to change her mind when confronted with new evidence. "She's a person who in the end accepts science."

Soon after the launch of *Down to Earth*, the CSE embarked on its anti-pollution campaign in Delhi and on one of its other top priorities — water. The organization helped communities use traditional know-how to collect and store rainwater. "When we began advocating this solution in the mid-1990s, people laughed at us, basically said we were idiots, that this was not a solution for the scale of India," says Narain. "Today, nobody will argue that with you."

Passionate about every aspect of her work, Narain is harder to pin down on personal questions. On the sunny day I visit the CSE offices, Narain is rushing about barefoot in her office, laughingly deflecting the conversation from her personal life. "I'm the world's most boring person. I come to work, I go to sleep, I know nothing else," she says.

In addition to answering my questions, Narain juggles other visitors, including local student journalists, frequent updates on the lawsuits the CSE is embroiled in, a television interview, and a domestic crisis precipitated by a fallen tree in front of her house.

Narain comes from one of Delhi's illustrious families: her grandfather was an eminent journalist and his brother a gandhian freedom fighter. Her father, also a young freedom fighter, died when she was eight, leaving her mother to bring up Narain and her three sisters.

Narain joined the CSE straight after school and pitched into whatever tasks were most urgent. She says Agarwal was a tough boss — she is reputed to be one herself — tearing up copy until the writers got it just right.

Agarwal was known for his intelligence and integrity, but many say privately that Narain can be more open-minded and has given the CSE a friendlier profile. Some even suggest she is too soft on the government — and too cosy with government insiders. Among Narain's friends, for instance, is prime minister Manmohan Singh's daughter. Such personal attacks offend Narain: "Who cares if I am well connected? I haven't gone to Manmohan Singh to ask for a job."

Narain may flinch at personal attacks but remains undeterred in her work. When I met her, she had just returned from Ahmedabad, the second city where the CSE has successfully pushed for a switch to compressed natural gas. "We've been fighting for two years to clean Ahmedabad's air," she says. "Yesterday, the air in Ahmedabad was clean. It's worth it." ■

B. MATHUR/REUTERS/CORBIS

The Centre for Science and Environment forced Delhi's public transport to switch from diesel to gas.

Apoorva Mandavilli is Senior News Editor for *Nature Medicine*.

See Editorial, page 683.

US air pollution is harmful and fine particles can kill

SIR — Your Special Report on air pollution in the United States ("The politics of breathing" *Nature* **444,** 248–249; 2006) has been criticized for accepting the conclusions of "two large, well-respected epidemiological studies" that long-term exposure to fine particles in the air is associated with increased mortality. Suresh Moolgavkar, in Correspondence ("Pollution analysis flawed by statistical model" *Nature* **445,** 21; 2007), says that there is "by no means universal agreement among scientists that air pollution at contemporary US levels affects human health".

That statement is probably correct. When do we ever see universal agreement about anything? However, he is not correct when he states that "joint pollutant analyses — with sulphur dioxide and either sulphates or fine particles both included in the statistical models — show that sulphur dioxide is associated with mortality; fine particles are not".

Moolgavkar refers to D. Krewski and colleagues' *Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality* (Health Effects Institute, 2000). But this reanalysis clearly supports the view, which I strongly hold, not only that air pollution at contemporary levels adversely affects health, but also that fine particles are most definitely associated with mortality. For example, on page 31 Krewski *et al.* state: "Nonetheless, both fine particles and sulphate continued to demonstrate a positive association with mortality even after adjustment for the effects of sulfur dioxide in our spatial regression analyses".

Steve Moorhouse
Environmental Health Division, Milton Keynes Council, Civic Offices, Milton Keynes MK9 3HH, UK

Quality evaluation needs some better quality tools

SIR — In their Commentary "Measures for measures" (*Nature* **444,** 1003–1004; 2006), Sune Lehmann and colleagues report that some widely used research 'quality' evaluation tools based on citation indices are unreliable, because these measures do not minimize statistical uncertainty. This highly interesting study shows, disturbingly, how little evaluation measures have been scrutinized by users so far. However, I would like to point out one additional, major issue.

The quality of publications cannot be measured using citation indices, whatever their statistical reliability. Instead, measures based on received citations assess the visibility of publications, authors or journals, and this may not necessarily correlate well with quality. For instance, if authors criticize a study for using unconvincing methods or for drawing the wrong conclusions, they will still be citing that study, thereby improving its citation record.

Also, when faced with many journals' instructions to limit the number of references, authors have to choose between references of equal quality. The choice of the citing authors will then inevitably rely on other criteria, such as convenience of access; also, geopolitical factors may play a role, as the analysis of citation flow between countries reveals distinct country clusters of citation preferences (A. Schubert & W. Glänzel *Scientometr.* **69,** 409–428; 2006). Finally, in some cases, authors who cite a paper may not even have read it (M. V. Simkin & V. P. Roychowdhury *Complex Syst.* **14,** 269–274; 2003).

In my view, there are promising models that could be developed into more targeted quality measures than currently used citation-based indices.

One example is that of the Internet-based commercial bookselling companies that have developed platforms where users — readers — evaluate the quality of publications; here, each evaluation itself can also be evaluated by other readers. The journal *Behavioral and Brain Sciences* publishes extensive open peer commentary alongside a target article and the authors' response to the comments. A combination of both models, asking peers to briefly evaluate a scientific publication according to several quality parameters (such as quality of methods, completeness, innovative potential or potential of generating beneficial effects to society), might serve better to draw a reliable and targeted picture of research quality than measures that only count citations.

Thomas F. Döring
Division of Biology, Imperial College London, Silwood Park campus, Ascot, Berkshire SL5 7PY, UK

Readers are welcome to comment on this Correspondence and the Commentary at Nautilus, our blog for authors: http://blogs.nature.com/nautilus/2007/01/post_1.html — Editor, *Nature*

Fossils: professionals and amateurs can cooperate

SIR — The statement that "commercial fossil trading in the United States has its roots in the 1960s" in your News story "Palaeontology journal will 'fuel black market'" (*Nature* **445,** 234–235; 2007) is misleading. Parts of the US National Museum of Natural History's collection of invertebrate fossils were bought in the late nineteenth century from amateur collectors. The collector's payment sometimes included a position as a research scientist in the museum as well as money for the material supplied.

The relationship between amateur and professional palaeontologists in the United States has been, and continues to be, a strong and positive one. It is unfortunate that the unscrupulous behaviour of individuals is sometimes misconstrued as revealing a divide between these communities.

Nigel Hughes
Department of Earth Sciences, University of California, Riverside, California 92521, USA

Ignore the spurious claims of private fossil-hoarders

SIR — Your News story about an 'amateur' journal gives some of the reasons why details of privately held fossils should not be published ("Palaeontology journal will 'fuel black market'" *Nature* **445,** 234–235; 2007). I would add that fossils in private collections are not essential to science.

Some very naive students of palaeontology think that finding new fossils is the only way to make an advance in the field. However, palaeontology is only one of many sciences studying evolution. The meaning of fossils — as a whole and as individual specimens — is linked to the other sciences and changes with the progress made in them.

Fossils found in the time of Georges Cuvier, for example, are no longer interpreted in the way that this remarkable scientist interpreted them 200 years ago. They do not have the same meaning. But as long as they remain available, they can be studied anew in the light of new knowledge. Our knowledge of evolution has changed since Cuvier's time: the geological context is better understood, the discovery of other fossils allows a better understanding of their relationships and new techniques give rise to new observations. So the fossils found in his time can be studied productively time and time again.

Fossils are eternal, and this is the main reason why we have public museums. In these institutions, all fossil specimens must be available for research and presentation to the public. We do not need fossils kept in private collections and available only to a few favoured people.

Jean-Louis Hartenberger
Laboratoire de Paléontologie cc 64, Institut des Sciences de l'Évolution, Université Montpellier 2, 34095 Montpellier cedex 5, France

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. Published contributions are edited.

BOOKS & ARTS

Choose your own reward

Does human creativity stem from a process that turns arbitrary ideas into goals like food and sex?

Why Choose This Book? How We Make Decisions
by Read Montague
Dutton: 2006. 331 pp. $24.95, £15.99

Andy Clark

Why don't our brains get as hot as the processors in our personal computers? And what does that tell us about biological computation, the nature of choice, the value of value signals, and the power of ideas? These questions may seem rather disparate, but Read Montague's provocative and accessible treatment of them in *Why Choose This Book?* displays a deep and unexpected unity. The story goes something like this. Biological computation is constrained to be efficient: to confer close to maximal computational power for close to minimal energetic expenditure. The outward sign of this efficiency is said to be the mere warmth of the human brain compared with the searing (wasted) heat of those computer processors. To be efficient, biological computations need to be equipped, so the argument goes, with some kind of measure of their own value, in relation to the in-built goals of maintaining life and reproductive success. Such measures allow the system to expend energy only on those computations that matter most. How such a measure might work remains problematic, but once it is in place, general principles of thrifty processing, such as the slow use of power, compression of data, conservation of wiring, and frugal use of bandwidth and communication, are all recruited to the mix.

But it is the goals and value signals that play the lead role in Montague's story. He introduces us to the guiding principle that will link efficiency to choice and to the power (and pathologies) of ideas. That principle, familiar enough in cognitive scientific circles but here tweaked and nuanced in novel and potentially transformative ways, goes by the unpromising name of 'reinforcement learning'. In reinforcement learning, goal states are approached by sensitivity to signals that predict rewards (the attainment of goals). But the system is not simply hardwired to regard only some fixed set of signals as reward signals; it obtains flexibility by learning associations between experienced signals and temporally removed (but consistently associated) rewards. Past experience of what signal leads to what reward is thus combined with present feedback (what's here now, and what is it worth?) to generate choices that (ideally) maximize total future reward.

R. BAMBER/ALAMY

Neat trick: activities such as gambling and card games can supplant the basic rewards of food or sex.

Moving all this along is the 'reward-prediction error signal', which carries information about how well the actual rewards tally with the predicted rewards. When the actual reward exceeds the predicted one, it makes sense to upgrade the stored value of the states that predicted the unexpectedly greater reward. In the brain, dopamine neurons provide at least one means of mechanistically encoding just such a reward-prediction error signal. Bursts of dopaminergic activity result when the reward exceeds the predicted reward; pauses in activity mean that the reward falls short of the predicted reward; and unchanged activity means the reward was as expected.

But what counts as a reward anyway? The most obvious rewards are the basic biological achievements of life maintenance (such as the ingestion of a tasty and nourishing morsel) and reproduction (or rather its precursor, sexual intercourse). Montague is motivated, however, by a strong desire to unravel the mechanistic underpinnings of what he describes as a uniquely human 'superpower': the capacity to make choices that seem to value biologically arbitrary objects, achievements and actions. Examples of such biologically arbitrary goal states mentioned in the text include solving Fermat's last theorem and committing group suicide in the belief that a spaceship hidden in a comet's tail will then take you to 'the next level'. What makes all this possible, in Montague's model, is the capacity of ideas themselves to act as reward signals, hijacking the prediction-error systems implemented by dopamine neurons in the brain. When this happens, the dopamine outputs start to act as error signals that encourage the rest of the brain to learn and to make decisions in ways that increase the chances of acquiring some biologically arbitrary reward.

Given the potentially biologically catastrophic consequences of such re-tooling of mere thoughts as rewards, Montague suggests that powerful filtering processes control what gets into the reward slot. But such processes can be fooled — in ways that the book describes in compelling and often sinister detail — by damage, by drug abuse, and perhaps even by some forms of advertising and branding (brands are just cues that predict rewards). Montague's proposal is that biologically arbitrary goals can somehow plug into a kind of 'special status reward socket', and thus become a basic, primary reward, like food or sex. He does not claim that these ideas become associated, either directly or indirectly, with food or sex; rather,

they plug directly into the 'socket' normally occupied only by the most basic high-status rewards. If we humans have indeed learnt such a powerful trick, it is no surprise that it fuels so much that is both good (creative and expansive) and ill (pathological and restrictive) in our species. Montague begins by laying out this possibility, then follows it deep into the fascinating territories of creative thought, addiction, obsessive–compulsive disorder, Parkinson's disease, and then on to the psychosocial realms of trust and regret.

Despite its attractions, there are some important mechanistic gaps in the story, as Montague acknowledges. For example, it isn't clear why or how one idea might win out over another in the bid to occupy a high-status reward socket, or how the occupation itself is accomplished. Nor is it really clear when such occupation should be deemed pathological rather than creative. I was also left wondering whether the basic idea of each symbol and each computation carrying its own value 'tag' — the difference, Montague argues, between standard computational models using 'meaningless symbols' and the hyper-efficient, value-rich computations said to be characteristic of biological nervous systems — is sufficiently clear and workable. Exactly how do these computation-value or symbol-value pairs work, and how do they transform mere symbol processing into meaning? Do they compose? Two computations whose individuals values are low might together constitute a complex computation whose value to the organism is high, but Montague suggests no way of systematically predicting such combined values from the values assigned to the parts.

Perhaps I am missing something, but it repeatedly struck me that Montague's overall vision is both rather more radical, and rather less mechanistically clear, than his book suggests. The prospective reader should be aware that the story on offer actually departs quite a long way from the basic computational theory of the mind. It builds in value and computation right down to the cellular level, and (more generally) systematically blurs the usual distinctions between life, mind and information processing. This blurring is evident, for example, in the puzzling idea that each individual neuron, in the quest for efficient interneural communication, might need to contain up to 100 million 'dynamic models' of other neurons and neuronal subsystems.

These are not really complaints, however. The book spans several seldom-bridged worlds, from neuroscience to psychiatry, economics and social psychology, and does so with wit, precision and elegance. It succeeds in many of its goals. Above all, it left me feeling I had actually learnt something about myself: a thinking, feeling, choosing, yet painfully vulnerable chemically modulated learning machine. ■

Andy Clark is at the School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh EH8 9JX, UK.

NASA KSC

James Webb (second left, with President John F. Kennedy) led NASA's Apollo space programme.

Shooting for the Moon

The Man Who Ran the Moon: James Webb, JFK, and the Secret History of Project Apollo
by Piers Bizony
Thunder's Mouth/Icon Books: 2006.
256 pp. $24.95/£16.99

Alex Roland

James Webb was the antithesis of the enterprise that made him famous. Administrator of NASA during the Apollo space programme, Webb commanded a sprawling, anonymous team of 400,000 workers. Yet he himself was colourful, singular, eccentric even — a politician and an individualist leading an army of technicians. Apollo was 'big science', what Webb and others called 'large-scale technology'. It dwarfed the Manhattan Project, for example, in cost, size and complexity. In contrast to the Wright brothers, who achieved atmospheric flight using equipment from their bicycle shop, the flight to the Moon integrated civil servants and industry contractors in a nationwide web of activity. They built unprecedented facilities such as the launch centre at Cape Canaveral and the manned-spacecraft centre at Houston Texas, and dispensed some $25 billion (worth several times as much today) over a decade or more. Such undertakings demand the subordination of the individual to the cause. Webb was certainly a team player, but he was also the pilot who steered his own course and imposed his personality on Apollo.

Webb's background and the purportedly "secret history" revealed in *The Man Who Ran the Moon* by Piers Bizony have long been well known. An accomplished Democratic operative and former budget director for President Harry Truman, Webb accepted the NASA position after as many as 17 other prospects had turned it down. He ran NASA (not the Moon) like a chief executive, handling the politics of external relations and leaving the technical management of the programme to trusted subordinates. He survived power struggles, budget battles and conflicts with two presidents, John F. Kennedy and Lyndon B. Johnson, over maintaining an overall space programme balanced between manned and unmanned activities. And he won the right to distribute some of NASA's money to support socially constructive programmes such as university development. Smart, energetic, gregarious and iconoclastic — his official limousine was a Checker cab — he cut a colourful swath through the bureaucratic maze of Washington.

Then came the Apollo 204 tragedy. On 27 January 1967, three astronauts died in a gruesome fire during a routine ground test of the Moon capsule at Cape Canaveral. In the ensuing investigation, an internal NASA report came to light that criticized the capsule manufacturer, North American Aviation (NAA), for shoddy work. Caught out on the witness stand without foreknowledge of the report, Webb began to lose control of events. Worse still, he appeared to misrepresent the fact that he and his leading deputies had overturned the recommendation of their own review panel and awarded NAA the contract in the first place. This revelation tarred him with the brush of the Bobby Baker scandal, a rat's nest of lobbying, bribes and even organized crime.

Webb survived the hearings but his reputation was damaged and his power diminished. New executives were brought in to restore

technical order in the programme, and Webb had less authority to impose his will on them. His determination to maintain a balanced programme and channel NASA funding into socially beneficial schemes was being overtaken by the budget and political crises of the Vietnam war. When he mentioned retirement to President Johnson in 1968, the president hastily called a press conference and practically pushed Webb out of the door. Both men were to attend the launch of Apollo 11 the next summer, but took less joy in the achievement than they might have otherwise.

Bizony tells this familiar story clearly and engagingly. To the existing literature he adds some interviews, primarily with Robert Seamans, deputy administrator of NASA under Webb. He quotes extensively from these sources, occasionally without making clear who is speaking. The result is 'Webb light', a fast-paced, breezy account weak on substance and contextualization. The book climaxes with the Apollo 204 crisis, followed by an impressionistic survey of NASA history since Webb.

The book is reasonably accurate and the undocumented opinions are plausible, but the account is unreliable on the details and silent on the complexity of Webb and the times in which he operated. Similarly, Webb's system of 'management by exception' is not discussed at all. Even so, it is an entertaining introduction to Webb, but it should be supplemented with W. Henry Lambright's *Powering Apollo* (Johns Hopkins University Press, 1995), Arnold Levine's *Managing NASA in the Apollo Era* (NASA, 1982) and Webb's own *Space Age Management* (McGraw-Hill, 1969). These books offer further insight into whether complex scientific and engineering projects on the scale of Apollo, with all their conflicting political, budgetary and technical demands, are manageable in any sense that Webb would have understood. ■

Alex Roland is in the Department of History, Duke University, Durham, North Carolina 27708, USA.

Opposites attract

Measuring the World
by Daniel Kehlmann,
transl. by Carol Brown Janeway
Pantheon: 2006. 259 pp. $23. To be published in the UK by Quercus in April.

John Whitfield

Quite often, it strikes me that being a scientist is an odd way to spend your time. We all ask the same questions. Where do I come from? Where am I going? What does it all mean? Yet few — and only relatively recently — have chosen the scientific method as the means to answer them. And for those who have, many of their answers seem as impenetrable and marginal as avant-garde poetry or 'squeaky gate' music.

Daniel Kehlmann's neat novel *Measuring the World*, a bestseller in Germany last year under the title *Die Vermessung die Welt* and now translated into English, provoked these thoughts once more. The book is set in the late eighteenth and early nineteenth centuries, when the structures of science, and the job of being a scientist, began to take on something like their present form. It weaves together the stories of two of the giants of the time: the mathematician Carl Friedrich Gauss and the explorer, geographer and biologist Alexander von Humboldt.

Kehlmann deploys the two men as archetypal and opposite examples of how to be a scientist. The core of Humboldt's story is his five-year journey to the Americas, which made him famous and had a huge influence on nineteenth-century naturalist travellers including Charles Darwin and Alfred Russel Wallace. But the journey makes no apparent impression on Kehlmann's Humboldt. He is the embodiment of cold rationality, a Gradgrind who, lacking any personality or inner life, builds one out of facts and measurements. He'd rather stare down his sextant than look at a solar eclipse, and rather study a woman's lice than have sex with her. He chases up rivers and mountains, oblivious to hardship, with French botanist Aimé Bonpland as his Sancho Panza.

Gauss, on the other hand, hates going anywhere. But then, he doesn't need to — from childhood, revelation comes to him, in an unbidden stream of mathematical genius. He sees science as "a man alone at a desk, a sheet of paper in front of him". This is also novel writing, so perhaps it is not surprising that Kehlmann makes Gauss the more sympathetic and, despite his freakish abilities, the more human character. He worships his mum, falls in love, visits prostitutes, and has children who disappoint him.

Humboldt is a cipher. This also has the effect of making Gauss's way of doing science seem more noble and authentic than Humboldt's. It isn't, but this is a neat twist, as mathematicians are usually the ones portrayed as weirdos.

It would be just as silly to complain that Gauss and Humboldt probably weren't much like this as it would be to object to Peter Shaffer's play *Amadeus* on the grounds that Salieri probably didn't aim to bump off Mozart. I will, however, make one point in Humboldt's defence. Kehlmann is truthful to the facts of his biography, and Humboldt was an enigmatic man, who tried to destroy documents pertaining to his early life, and who might have substituted work for emotional fulfilment. (There has been speculation, to which Kehlmann briefly alludes, that this is because Humboldt was homosexual.) But he knew how to do the right thing. Bonpland returned south to America but was caught in disputed border territory and imprisoned. Kehlmann's Humboldt wrings his hands; the real Humboldt, in contrast, sold his world-class collection of plant specimens to provide his friend with financial support.

Kehlmann skilfully stops *Measuring the World* becoming a highbrow tale of nutty professors. For a start, his professors are more melancholic than nutty. Gauss's prodigious abilities — and his decision to be true to them, even at the cost of his own and others' happiness — cut him off from people, and everyone else's stupidity depresses him. Humboldt's political, administrative and official duties gradually overwhelm his opportunities to take measurements, and in old age he reprises his American journey in Russia, as farce. Each learns that no degree of cleverness or immersion in science grants immunity from, or even helps much with, the messy business of life and death.

Kehlmann also avoids naffness by telling

THE ART ARCHIVE/M. CHARVET

NATIONALGALERIE, BERLIN/BRIDGEMAN ART LIBRARY

Carl Friedrich Gauss (left) and Alexander von Humboldt had very different views of how science should be done.

his story in a relentless deadpan, which is at first alienating but then gets under the skin. As the story develops, your sympathy for the two men grows, as their own does for each other. One of the things they agree on, for example, is the deplorability of "novels that wandered off into lying fables because the author tied his fake inventions to the names of real historical personages".

Kehlmann, then, does a good job of capturing the strangeness and comedy of science, as well as the powerful sense of futility that can afflict researchers from time to time. But he doesn't get near to explaining why, despite its oddness, science provides such powerful and beautiful answers to our questions, or why mathematics has such an uncanny power to provide these answers. Nor do we get any idea why a few people, such as Gauss, have mathematical abilities that seem supernatural to the rest of us, or why others, such as Humboldt, are willing to give up their fortunes, comforts and sometimes lives to see, and measure, what's over the horizon. Quite right too, I'm tempted to think — where would the fun be in knowing that? ■

John Whitfield lives in London and is the author of *In the Beat of a Heart: Life, Energy, and the Unity of Nature* (Joseph Henry Press).
www.inthebeatofaheart.com

SCIENCE IN CULTURE

The molecular landscape

Lucia Covi uses modern microscopy to highlight the world at the nanoscale.

Martin Kemp

"We are 'connoisseurs of chaos', patterners. So we look for resemblances to things in our experience... The gold tip is a digital Tower of Babel, or a wedding cake. And a fractal set, and the electron microscope image I once saw of a small worm's mouth."

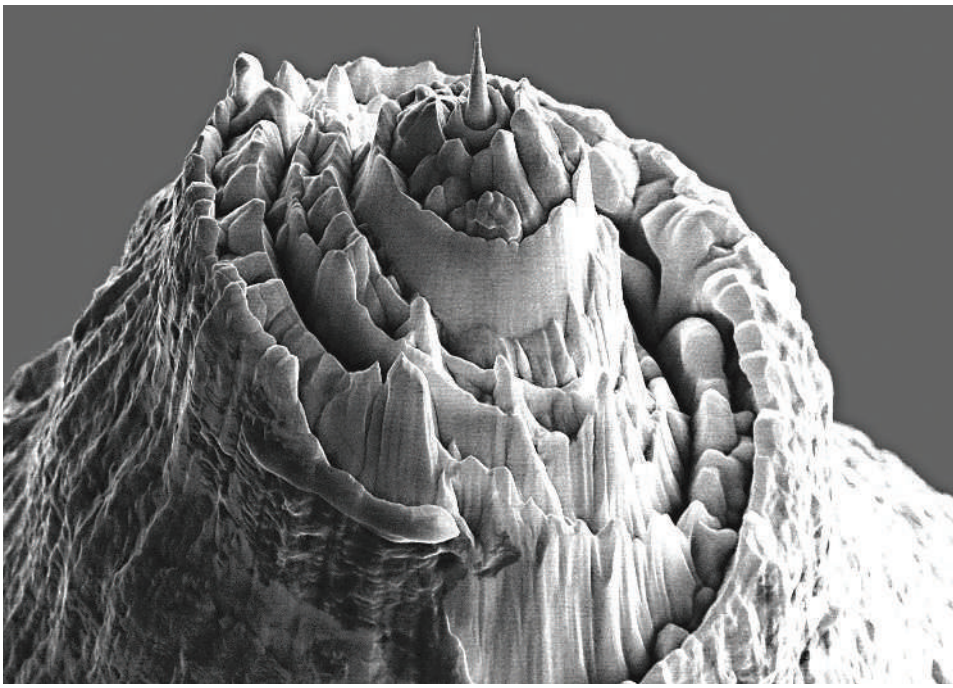
So says Roald Hoffmann, winner of the 1981 Nobel Prize in Chemistry, in his introduction to Lucia Covi's book *Blow Up: Images from the Nanoworld* (Damiani Editore, €26; www.damianieditore.it). Hoffmann's brief essay should be compulsory reading for anyone involved with machine-generated images in either science or art.

Covi is a Milanese photographer who has worked with Elisa Molinari and her colleagues at the Italian National Research Center on nanoStructures and bioSystems at Surfaces in Modena. Together they have made extraordinary structures visible at the scale of millionths of millimetres. Covi's book was published to mark an exhibition previewed at the 2006 Genoa Science Festival and now on show in Modena (www.s3.infm.it/blowup). It also stands on its own as a visually and intellectually stimulating panorama of images from the strange yet somehow familiar nano-landscapes of modern microscopy.

Covi has turned scientific data into 'photographs' to magnify their visual impact. This often involves limiting the propensity of scientists to render their computer images in garish colours, a cacophony of metallic hues and tones. To make the images appear convincing and highlight their beauty, all the visual effects need to be internally consistent. As Hoffmann declares: "Differences in surface texture, in smoothness and roughness matter. They are compared in our brain with memories of tangible objects." Internal visual consistency is an incredibly subtle matter.

The image of the gold tip (shown here), specifically mentioned by Hoffmann and used on the cover of the book, was captured by a scanning electron microscope focused on the tip of a probe used in scanning near-field optical microscopy — one microscope in effect scrutinizing another.

Just like Robert Hooke, when he described the wondrous little 'engines' and landscapes he witnessed when compiling his *Micrographia* in 1665, we automatically draw perceptual parallels with familiar objects when we see new structures.



Making a point: a gold tip used in microscopy resembles Brueghel's Tower of Babel.

S3 NATIONAL RESEARCH CENTER (INFM-CNR)

Now, in this age of fractal landscapes in science fiction and animated films, my first thought when seeing the gold tip is of a fantasy castle constructed on a conical mountain-top. Or, to pick up Hoffmann's more erudite, historical parallel, the Tower of Babel as characterized in Pieter Brueghel's amazing sixteenth-century painting.

The gold tip is an artefact of the nano-sculptor's craft. It was sculpted by milling with a focused ion beam, a top-down process. Other images in the book show structures that have self-assembled spontaneously, in a bottom-up manner.

At this scale — as perhaps at every level of the structural organization of the material world — the basic building blocks aggregate and form into morphologies that are recognizably regular yet irreducibly individual and unpredictable.

My one quarrel with Hoffmann is when he declares: "No one is born with a feeling for harmonious arrangement." My conviction is that our systems of perception and cognition are profoundly endowed with an innate propensity for discerning levels of order and disorder, which I have termed 'structural intuition'. The images in Covi's book present a veritable field day for the exercising of this propensity — whether we think it is taught or innate, or a compound of both.

Martin Kemp is professor of the history of art at the University of Oxford, Oxford OX1 1PT, UK. His new book, *Seen | Unseen*, is published by Oxford University Press.

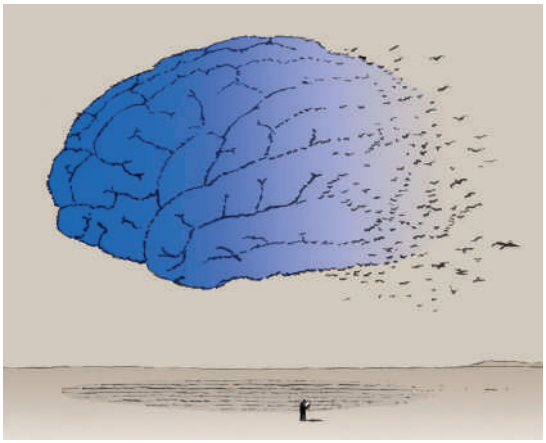
Collective minds

By tapping into social cues, individuals in a group may gain access to higher-order computational capacities that mirror the group's responses to its environment.

Iain Couzin

In 1905 the field naturalist Edmund Selous, a confirmed Darwinian and meticulous observer of bird behaviour, wrote of his wonderment when observing tens of thousands of starlings coming together to roost: "they circle; now dense like a polished roof, now disseminated like the meshes of some vast all-heaven-sweeping net...wheeling, rending, darting…a madness in the sky".

Throughout his life Selous struggled to explain the remarkable synchrony and coherence of motion during flocking, and he concluded that somehow a connectivity of individual minds and transference of thoughts must underlie such behaviour. "They must think collectively, all at the same time, or at least in streaks or patches — a square yard or so of an idea, a flash out of so many brains".



J. KAPUSTA/IMAGES.COM

We now know that such synchronized group behaviour is mediated through sensory modalities such as vision, sound, pressure and odour detection. Individuals tend to maintain a personal space by avoiding those too close to themselves; group cohesion results from a longer-range attraction to others; and animals often align their direction of travel with that of nearby neighbours. These responses can account for many of the group structures we see in nature, including insect swarms and the dramatic vortex-like mills formed by some species of fish and bat. By adjusting their motion in response to that of near neighbours, individuals in groups both generate, and are influenced by, their social context — there is no centralized controller.

But when observing a starling flock, or a sweeping, twisting school of tiny silvered fish, I often think of Selous's concept of a collective mind. It is perhaps too easy to disregard his vision, based as it was on a Victorian fascination with telepathy. Indeed, his rich descriptions capture the essence of something more, something we still know very little about: how social interactions affect the way animals within highly coordinated groups acquire and process information.

For individuals within groups, survival can depend critically on how local behavioural rules scale to collective properties. Pertinent information, such as the location of resources or predators, may often be detected by only a relatively small proportion of group members due to limitations in individual sensory capabilities, often further restricted by crowding. Close behavioural coupling among near neighbours, however, allows a localized change in direction to be amplified, creating a rapidly growing and propagating wave of turning across the group. This positive feedback results from the ability of individuals to influence and be influenced by others, and allows them to experience an 'effective range' of perception much larger than their actual sensory range.

The scaling from actual to effective sensory range is non-linear, however. It is hard for groups to remain cohesive and for information to spread if individuals respond only to others very close to themselves. As sensory range is increased, a response to a greater number of neighbours increases cohesion and allows effective long-range transfer of directional information. If this range expands further still, groups that form are highly cohesive but individuals may get misdirected, as the motion of distant individuals is less likely to encode relevant information about localized stimuli.

Individuals within groups may modify their interactions in a context-dependent way. Under threat of attack, for example, individuals often align more strongly with one another, heightening collective sensitivity to weak or ambiguous environmental stimuli, and so increasing the 'system gain'. However, amplification can occur in response to random fluctuations, creating false alarms that can be costly.

Under different circumstances individuals may adopt behaviour that facilitates collective damping of local fluctuations. During long-distance migration, for example, animals are often faced with the challenge of navigating up noisy and weak thermal or resource gradients. Local variability makes this task difficult, or even impossible, for individuals in isolation. But coherent social interactions can allow groups to function like an integrated self-organizing array of sensors, again increasing effective perceptual range. As long as interactions are sufficiently sensitive to ensure cohesion, but not too sensitive to local fluctuations and individual error, individuals can effectively respond to the weak long-range gradient.

We are beginning to comprehend more fully how individuals in groups can gain access to higher-order collective computational capabilities such as the simultaneous acquisition and processing of information from widely distributed sources. Group members may come to a consensus not only about where to travel but also about what local rules to use. Thus, like the brain, groups may adapt to compute 'the right thing' in different contexts, matching their collective information-strategy with the statistical properties of their environment.

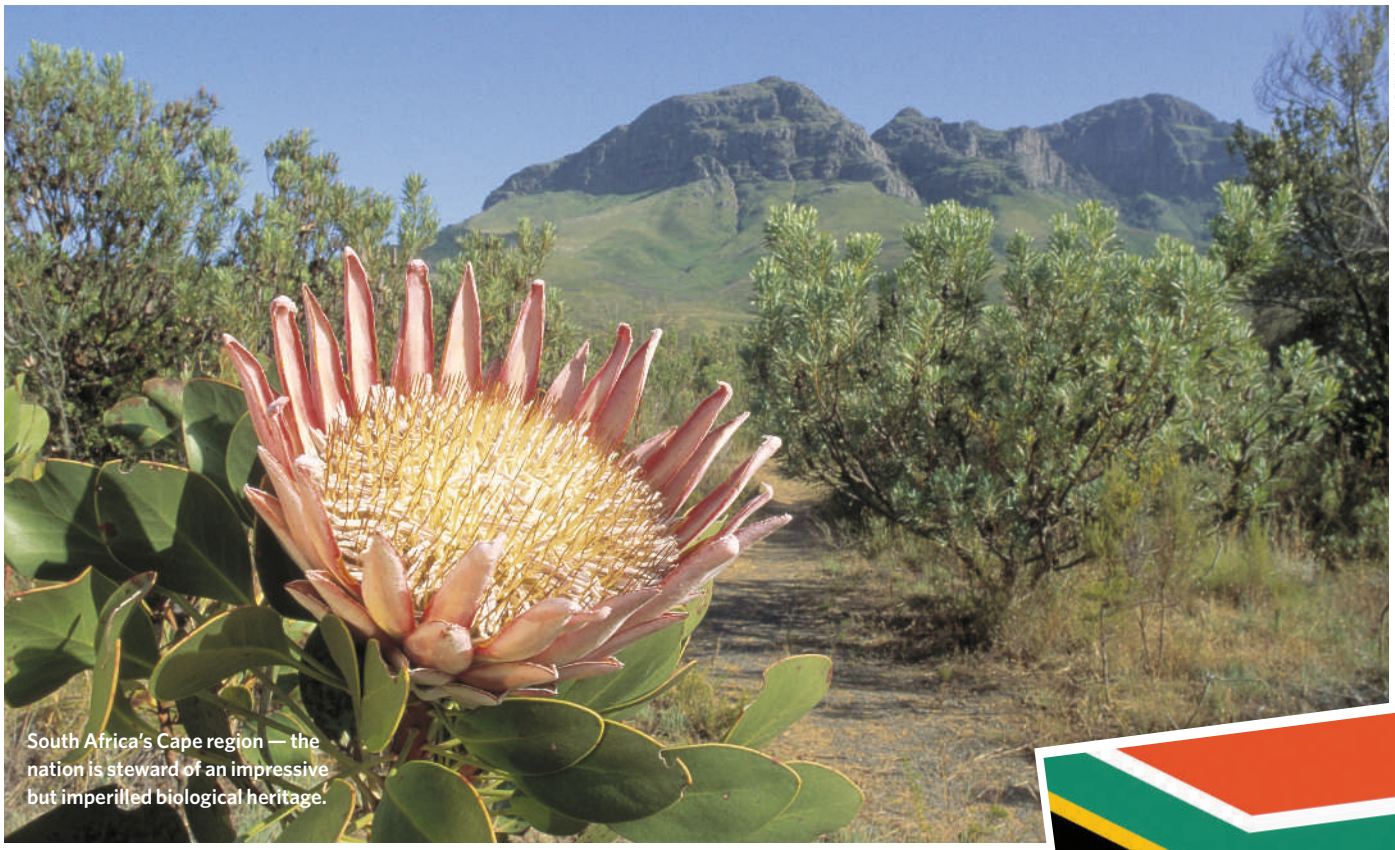
Selous wrote in despair of his contemporaries' lack of interest in flocking: "If there really were anything extraordinary in the collective movements of birds…they would have been much discussed and much wondered at". But today there is a rapidly expanding and vibrant community of biologists, engineers, mathematicians and physicists for whom flocking serves as inspiration. Such group behaviour holds clues about the evolution of sociality, and also for the development of novel technological solutions, from autonomous swarms of exploratory robots to flocks of communicating software agents that help each other to navigate through complex and unpredictable data environments. ■

Iain Couzin is in the Department of Zoology, University of Oxford, Oxford OX1 3PS, UK.

FURTHER READING
Conradt, L. & Roper, T. J. *Trends Ecol. Evol.* **20,** 449–456 (2005).
Couzin, I. D. & Krause, J. *Adv. Study Behav.* **32,** 1–75 (2003).
Couzin, I. D., Krause, J., Franks, N. R. & Levin, S. A. *Nature* **433,** 513–516 (2005).
Sumpter, D. J. T. *Phil. Trans. R. Soc. B* **361,** 5–22 (2006).

For other essays in this series, see http://nature.com/nature/focus/arts/connections/index.html

NEWS & VIEWS



N. J. DENNIS/NHPA

South Africa's Cape region — the nation is steward of an impressive but imperilled biological heritage.

CONSERVATION BIOLOGY

The diversity of biodiversity

Arne Ø. Mooers

Species richness is not the same as evolutionary richness. So which is the better measure for setting conservation priorities? The flora of the Cape of South Africa provides a test for that pressing question.

South Africa's distinctive flag symbolizes its diverse cultural heritage. The nation's biological heritage is no less impressive: whereas the British Isles is home to about 1,500 plant species, South Africa's Cape region houses more than 9,000 in one-third of the area, with perhaps the highest concentration of endangered plant species anywhere in the world[1]. On page 757 of this issue, Forest *et al.*[2] report how an evolutionary approach to plant conservation might lead to surprising choices among conservation areas. The study brings the concept of what might be called evolutionary heritage into sharp focus.

The species is the fundamental unit of biodiversity, and so the most common conservation approach is to concentrate resources in the most species-rich areas. This assumes that all species are equivalent, but species vary dramatically in their evolutionary isolation[3]. Most have many and similar brethren (there are 3,400 members of the rose family, for instance), whereas some have only distant relatives (the evergreen shrub *Amborella trichopoda* seems to have no close relative among the entire ensemble of flowering plants). Biodiversity is not just the number of species but also the differences between them (Fig. 1, overleaf), and it seems intuitively obvious that both number and difference should inform conservation decisions[4]. Indeed, algorithms that select areas to maximize the evolutionary divergence among species in a particular region are several decades old[5].

Intuition is fine, but theoretical work[6] has cast doubt on the need for such algorithmic approaches: species richness works well because areas with more species in a group generally also contain more of the group's total diversity. Given that measuring evolutionary divergence is costly, the cheaper approach of using species counts might be sufficient. Furthermore, whereas species number is a straightforward metric, it is not clear exactly what one would be conserving if one focused instead on evolutionary divergence. Forest *et al.*[2] confront both these issues with data.

Using an evolutionary tree linking the entire Cape flora and a comprehensive map of where each genus (species grouping) occurs naturally, Forest *et al.* compare genus richness with total evolutionary divergence (see Fig. 1 of the paper on page 758). The pattern they discover in South Africa is the sort that maximally decouples taxonomic richness (here, of genus rather than species) from total evolutionary divergence. In the western region, repeated evolutionary radiations — rapid bouts of evolution leading to many species — have led to mind-boggling numbers of taxa. The common

NATURE|Vol 445|15 February 2007

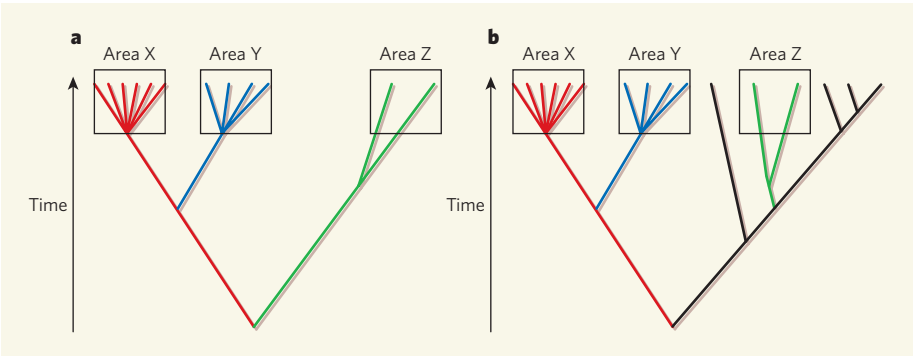


Figure 1 | Species diversity and evolutionary diversity. a, The evolutionary tree linking a set of hypothetical species found in three areas. Area X contains the most species and maximizes the total path along the evolutionary tree (in red). Area Y adds more species (blue), but area Z adds more unique evolution (green) to the tree linking the species in the three areas; so areas X and Z are priorities for preserving evolutionary heritage. **b**, The geographical scope is expanded and more of the tree is considered. If it turns out that species related to Z are already conserved (black), Y becomes the better choice for conservation. In terms of the Cape of South Africa, as studied by Forest *et al.*[2], areas X and Y represent the western region, and area Z represents the eastern region.

approach of maximizing the total number of conserved taxa would thus dictate choosing successive plots in the west. However, these genera are often each others' closest relatives. In the eastern region, the Cape vegetation meets and blends with another distinct floral assemblage (Maputoland–Pondoland–Albany), and plots are mixtures of more divergent genera. The maximum total evolutionary diversity measured in millions of elapsed years is retained with a mix of plots from both regions. The pattern of getting more divergence from fewer total species held even when Forest *et al.* estimated evolutionary divergence using a simpler (and so cheaper) taxonomy instead of their new evolutionary tree.

As the authors point out, the pattern predominant in the west (many closely related genera in the same area) is exactly what one would predict for areas that have seen repeated evolutionary radiations. Perhaps this is the pattern in other particularly species-rich areas of the world, for example in the intertidal areas of the northwest Pacific, or the tropical regions of the New World.

But how is preserving a larger slice of evolutionary diversity better than preserving more taxa? Forest *et al.*[2] test an argument based on future utility[7]. They first identify all the genera in their evolutionary tree that are already known to have economic uses somewhere in the world. These genera are widely dispersed across the tree, and they show that maximizing evolutionary divergence captures more of these genera than would random choice. In fact, because taxa in an area set aside for conservation will often be more closely related than they would be from random expectation, the authors' comparison is probably conservative. If so, then evolutionary divergence might be a far better approach than species richness if South Africa wishes to preserve the largest number of potentially useful taxa. Because the plots used here were so much larger (750 km^2) than most real preserves, we must wait for a more detailed data set to test this prediction directly.

The authors show both how and why a single country might best attend to its own evolutionary heritage. But the result presents a conservation conundrum. Of the bewilderingly many plant species in the Cape of South Africa, more than 6,000 are found in no other country in the world. The majority of these are in the western region. In contrast, the significant evolutionary diversity identified in the east is due in part to genera that are also found in neighbouring countries (Fig. 1). From the international perspective, perhaps South Africa should concentrate on its relatively closely related but exclusive western flora, and hope that Mozambique, Swaziland and so on will steward more divergent but shared taxa.

The issue of national interest and shared conservation responsibility is obviously not unique to plants, or to South Africa[8]. For example, the South Pacific island of New Caledonia is the sole home of both *A. trichopoda* and the endangered kagu, a bird whose closest relative is in an entirely different family. The iconic polar bear's fast-melting range spans half-a-dozen countries. Given differences in law, in priorities and in resources among nations, and given the current strain on the biosphere everywhere, the issue of who is responsible for what aspects of diversity is at once scientific, political and urgent. ■

Arne Ø. Mooers, currently a fellow of the Institute for Advanced Study, Berlin, Germany, is in the Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada.
e-mail: amooers@sfu.ca

1. Goldblatt, P. & Manning, J. C. *Cape Plants, a Conspectus of the Cape Flora in South Africa* (Natl Bot. Inst. S. Afr., Cape Town, 2000).
2. Forest, F. *et al. Nature* **445,** 757–760 (2007).
3. Purvis, A. & Hector, A. *Nature* **405,** 212–219 (2000).
4. Vane-Wright, R. I., Humphries, C. J. & Williams, P. H. *Biol. Conserv.* **55,** 235–254 (1991).
5. Faith, D. P. *Biol. Conserv.* **61,** 1–10 (1992).
6. Rodrigues, A. L., Brooks, T. M. & Gaston, K. J. in *Phylogeny and Conservation* (eds Purvis, A., Gittleman, J. L. & Brooks, T.) 101–119 (Cambridge Univ. Press, 2005).
7. Crozier, R. H. *Annu. Rev. Ecol. Syst.* **28,** 243–268 (1997).
8. Mooers, A. Ø., Heard, S. B. & Chrostowski, E. in *Phylogeny and Conservation* (eds Purvis, A., Gittleman, J. L. & Brooks, T.) 120–138 (Cambridge Univ. Press, 2005).

ANALYTICAL CHEMISTRY

Sense and versatility

A. Prasanna de Silva

Molecules that detect chemicals are the workhorses of analytical devices, but most recognize only one kind of target. A molecular sensor has now been devised that measures the concentrations of several metal ions.

What do the following things have in common: industrial effluents in a lake, signalling pathways in a cell and the blood status of a patient? The answer is that they can all be monitored by studying the concentrations of metal ions. There is a growing need for metal-ion detection, and one way of addressing this is to make sensors work harder. Writing in *Angewandte Chemie*, Schmittel and Lin[1] describe an impressive multitasking sensor — a molecule that binds to ions so that they can be interrogated with four different analytical techniques. This versatility allows the sensor to distinguish between various metal ions and to quantify their concentrations (Fig. 1).

Three of the detection techniques — referred to as channels — used by the authors have previously been individually established as ion-sensing methods. The first sensory channel detects the colour change of an indicator upon ion binding, a method that dates back to at least a century ago[2]. The second channel observes changes of luminescence. Many luminescent molecules emit a different intensity or frequency of light on binding to a metal compared with when they are metal-free, and these changes can be predictive of the type of metal that is bound[3]. The third channel senses changes in the electron density of molecules, brought about by the binding of metal ions. Such electrochemical behaviour is related to the reducing or oxidizing (redox) ability of the molecule[4], and is also a good sensory pathway.

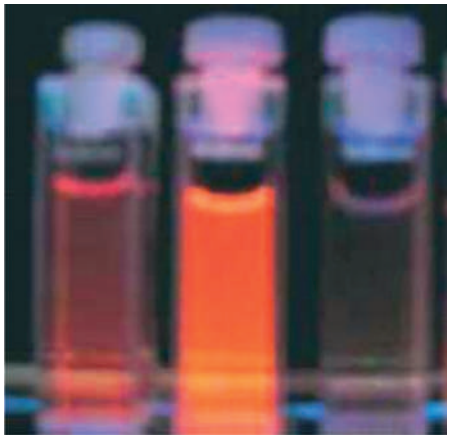


Figure 1 | A versatile molecular sensor. Schmittel and Lin[1] have prepared a luminescent compound that identifies different kinds of metal ion. The left-hand vial contains a solution of this compound. The intensity of its luminescence increases on the addition of lead ions (middle vial), but decreases in the presence of copper ions (right-hand vial). The ion sensor also has three other ways of distinguishing between ions. (Reproduced from ref. 1.)

The three channels mentioned above have previously been used together in various combinations[5], but Schmittel and Lin[1] now cleverly build on this foundation by adding the rarely used sensory technique of electrogenerated chemiluminescence[6]. This detects light emission from molecules following their electrochemical oxidation or reduction, and is therefore conceptually related to the other three methods.

The authors' sensor molecule is based on two components — an optical unit that can generate signals for all four channels, and a metal-binding receptor unit, which also contributes to redox activity. The optical unit is a luminescent ruthenium complex that hails from a family of compounds commonly used in some types of solar cell[7]. The receptor unit is an aza-crown ether[8] — a molecular loop that forms a three-dimensional crown shape in solution — that can encircle and ensnare metal ions.

The aza-crown ether is jostled by the chemical groups that surround the ruthenium centre, so that the receptor binds preferentially to different metals depending on the sensory technique used. This is because each method of detection deposits different net amounts of electric charge on the receptor. The sensor attempts to stabilize this charge by overlapping the molecular orbitals of the receptor with those of the optical unit, but the movement involved in this process is restricted by the jostling chemical groups. The amount of movement that occurs ultimately depends on the amount of charge on the receptor. As a result, the receptor is compressed to various extents depending on the sensory technique used, which in turn determines which metals fit into the receptor. Such multi-exploitation of a single kind of receptor is particularly clever — previously, sensors designed to handle several kinds of ion required a selective receptor for each of their targets[9].

As with any advance, this work suggests many possibilities for future studies. For most practical applications, the sensor must be adapted to measure ion concentrations in water, rather than in the organic solvents used here. The broad utility of this approach also needs to be demonstrated — do crowded receptors generally discriminate between different metals depending on the choice of detection method? And could other sensing channels be used, perhaps to detect targets other than metal ions?

Schmittel and Lin's work also shows how molecular devices[10] are increasingly flexing their minuscule muscles. The authors' molecule is single-handedly taking on the jobs of four sensors. This augurs well for the 'lab on a molecule' concept[9] — the idea that a set of tasks previously carried out by, for instance, a full clinical lab and a medical practitioner could instead be squeezed onto a single molecule. Such a possibility may seem implausible, but the concept stems from research in molecular logic and molecular computation[11], a field that itself was considered impractical not so long ago.

Schmittel and Lin's work belongs to the field of multi-mode transducers[5], molecular devices that convert chemical 'signals' into several other forms. An associated hot topic of research is the burgeoning area of array-based sensors, in which several molecular sensors are positioned on a single chip[12] or on the tip of an imaging fibre-optic[13]. The common thread of these approaches is that they detect many chemical targets; this might lead to increased efficiency in applications as diverse as cell physiology and environmental monitoring of industrial waste. Taken together, these developments send a clear signal that analytical chemistry is undergoing yet another reincarnation, reaffirming its reputation as a vibrant area of science and technology. ■

A. Prasanna de Silva is at the School of Chemistry and Chemical Engineering, Queen's University, Belfast BT9 5AG, UK.
e-mail: a.desilva@qub.ac.uk

1. Schmittel, M. & Lin, H.-W. *Angew. Chem. Int. Edn* **46,** 893–896 (2007).
2. Bishop, E. (ed.) *Indicators* (Pergamon, London, 1972).
3. de Silva, A. P. *et al. Chem. Rev.* **97,** 1515–1565 (1997).
4. Kaifer, A. E. & Gomez-Kaifer, M. *Supramolecular Electrochemistry* (Wiley-VCH, Weinheim, 1999).
5. Jimenez, D. *et al. Eur. J. Inorg. Chem.* 2393–2403 (2005).
6. Richter, M. M. *Chem. Rev.* **104,** 3003–3036 (2004).
7. Gratzel, M. *Inorg. Chem.* **44,** 6841–6851 (2005).
8. Izatt, R. M. *et al. Chem. Rev.* **85,** 271–339 (1985).
9. Magri, D. C., Brown, G. J., McClean, G. D. & de Silva, A. P. *J. Am. Chem. Soc.* **128,** 4950–4951 (2006).
10. Balzani, V., Venturi, M. & Credi, A. *Molecular Devices and Machines* (Wiley-VCH, Weinheim, 2003).
11. de Silva, A. P., Gunaratne, H. Q. N. & McCoy, C. P. *Nature* **364,** 42–44 (1993).
12. Wright, A. T. & Anslyn, E. V. *Chem. Soc. Rev.* **35,** 14–28 (2006).
13. Walt, D. R. *Science* **308,** 217–219 (2005).

50 & 100 YEARS AGO

50 YEARS AGO

A joint meeting of the Institution of Electrical Engineers, the American Institute of Electrical Engineers and the Engineering Institute of Canada was held on January 24, during which the lecture theatres of these three bodies in London, New York and Montreal were connected together by the trans-Atlantic telephone cable, which was opened for public service last September... At the end of the symposium the clarity of transmission over the cable was demonstrated when identical recordings of music were played in London and New York... Without the indicator provided in the lecture theatre in London, it would have required a very sensitive musical ear to determine the source of the music.
From *Nature* 16 February 1957.

100 YEARS AGO

Death has been very busy of late among the army of men of science, and nowhere has he been more active than in Russia, where within the space of a few weeks three of that country's foremost chemical philosophers—Beilstein, Mendeléeff, and Menschutkin—all men of front rank and of a worldwide reputation, have submitted themselves to the strict arrest of the fell sergeant... Our immediate concern is with the most distinguished of the eminent triumvirate—Dmitri Ivanovitsch Mendeléeff...he was a Siberian, born at Tobolsk on February 7th (N.S.), 1834. He died, therefore, within a week of his seventy-third birthday. He was the seventeenth and youngest child of Ivan Paolowitch Mendeléeff... The story of the rise and development of the Periodic Law is so well known that it is unnecessary now to dwell upon it. By a good fortune, which some may regard as evidence of predestination, Mendeléeff lived to see the verification of his predictions in the discovery, in rapid succession, of gallium, scandium and germanium; and no seer ever prophesied more truthfully.
From *Nature* 14 February 1907.

PLASMA PHYSICS

On the crest of a wake

Robert Bingham

What a conventional particle accelerator needs kilometres to achieve, a compact 'plasma wakefield' accelerator has just mastered in less than a metre. So is it adieu to the era of the gargantuan mega-accelerator?

Wakes — the areas of fluid turbulence most commonly seen around the pontoons of bridges and behind moving boats — have long excited human curiosity. Leonardo da Vinci carried out some of the first scientific investigations into wakes in the early sixteenth century by placing obstacles in fast-moving water (Fig. 1). During the 1830s, the Scottish naval engineer John Scott Russell noted that, at certain speeds, a boat travelling along a canal is actually accelerated by its own wake.

On page 741 of this issue, Blumenfeld *et al.*[1] take the exploitation of this wake effect to a new, higher-energy plane. They describe a technique for accelerating electrons in the wake of an ultra-relativistic electron beam propagating through an ionized gas (a plasma). Over a distance of less than a metre they succeed in doubling the energy of the electron beam at the Stanford Linear Accelerator Centre (SLAC) in California — which reaches 42 gigaelectronvolts (GeV) only after passing 3 kilometres of conventional technology based on radio-frequency accelerator cavities.

The result clearly demonstrates that the ultra-high gradients of 'plasma wakefield accelerators', which have previously operated[2–7] on scales of millimetres to centimetres, can be extended, in a metre-scale plasma, towards the high-energy frontier. That frontier currently lies at about 115 GeV, the energy achieved by the 27-km-circumference Large Electron–Positron (LEP) collider at CERN, the European particle-physics lab near Geneva, in its final days in 2000 before it was dismantled to make way for CERN's new Large Hadron Collider. What's more, as the authors show, there is no fundamental barrier to extending high-gradient plasma accelerators to work over arbitrarily long distances.

In a conventional accelerator, the accelerating radio-frequency electric field is limited to the 'breakdown field' at which it begins to rip electrons from the surrounding metal (often superconducting) accelerator cavities. But a plasma is already broken down, which means that it can support electric fields orders of magnitude higher than can conventional accelerators. In the SLAC experiment[1], the factor is at least 1,000; correspondingly, the plasma is more than 1,000 times shorter for the same energy gain than the conventional part of the accelerator.

Plasma wakefields can be created by either a single, intense laser or by a particle beam. Last year, laser-driven wakefield accelerators were used to produce monoenergetic electron beams in the gigaelectronvolt energy range over a few centimetres[6]. Blumenfeld and colleagues' accelerator[1] is driven by the tightly packed relativistic electrons of the SLAC beam entering a metre-long column of lithium vapour (Fig. 2). The intense electric field of the beam immediately strips the electrons from the gas with such force that they are blown outwards in all directions, leaving behind the more massive lithium ions. To restore charge neutrality to the plasma, the displaced electrons snap back away from the forwards-moving pulse, overshooting their original positions.

This oscillation in the plasma's electron density creates an oscillating electric wakefield, much like the wake produced by the boat in John Scott Russell's experiments, that accelerates part of the beam that formed it. Although the core of the electron pulse loses energy in setting up the intense plasma wakefield, the wakefield accelerates a small number of the electrons from 42 to 85 GeV over a distance of just 85 centimetres. A previous experiment[5] had achieved an energy gain of 2.7 GeV in a 10-centimetre-long plasma of similar density.

The SLAC experiment clarifies that an electron pulse can indeed propagate stably in a dense plasma for long distances without breaking up. No transverse deflections — 'sloshing' or 'hosing'[8] caused by the focusing effect of the ion column left behind by the passage of the electron beam — were observed. Hosing instability amplifies any asymmetry in the head-to-tail alignment of the beam, making directing the beam very difficult in longer systems. The lack of hosing also reduces 'betatron' emission, which occurs when electrons off the beam axis are attracted by the ion column, and wiggle back and forth across the axis emitting X-rays[9].

The maximum energy achieved in Blumenfeld and colleagues' experiments is determined by the expansion and consequent erosion of the beam head. The foremost part of the ionizing electron beam is not subject to the transverse focusing forces of the trailing ion column; the electric field of these electrons thus falls below the threshold for the formation of the plasma. This limiting effect can be overcome by higher-quality beams, in which electron spreading occurs only over distances much greater than the length of the plasma column.

Although these results are encouraging, current accelerator technology based on radio-frequency cavities is very successful, and has 50 years of development behind it. Replacing existing technology, wholly or in part, with plasma-wakefield acceleration will require further long and sustained effort. Maintaining high accelerating gradients over long plasmas is just one of a number of steps that need to be achieved. Particle colliders also require beams with an energy spread much smaller than that achieved by Blumenfeld *et al.*[1]. One way of achieving such a 'monoenergetic' beam is to boost the energy of a distinct second bunch trailing the drive bunch that sets up the wakefield — the afterburner concept[10]. Such a concept could be incorporated into a hybrid machine in which current mature technology produces an energetic beam up to 100 GeV, and the plasma accelerator takes it up to energies in the teraelectronvolt range over several metres.

Another essential ingredient of a collider is a counter-propagating positron (anti-electron) beam. Positrons have been accelerated in the wakefields of lower-density

Figure 1 | Turbulent history. A rigid obstacle in flowing water creates wake turbulence, a fact noted and sketched by Leonardo da Vinci in 1509. The passage of a moving body through a static medium has the same effect, and the turbulence can be used to accelerate a following body. What is true for a boat in water also applies to electrons passing through a gas, a fact exploited by Blumenfeld *et al.*[1] in their high-energy plasma wakefield accelerator.

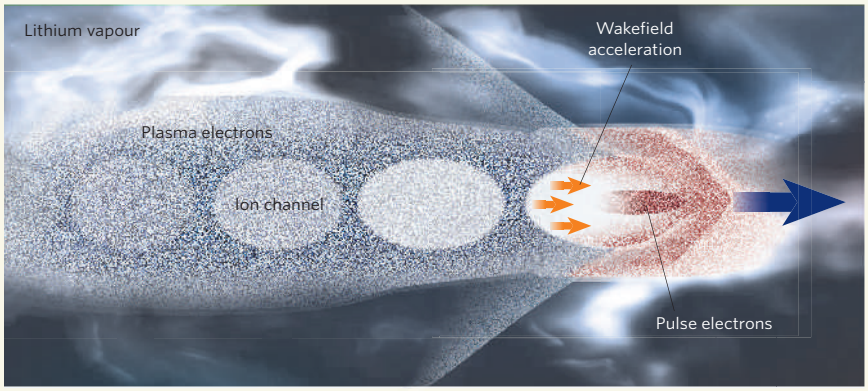


Figure 2 | Strip and surf. In Blumenfeld and colleagues' experiments[1], the 42-GeV Stanford Linear Accelerator electron beam ionizes lithium vapour with such force that it strips off all the atomic electrons, leaving only lithium ions behind. Most of the electrons lose energy in setting up the wakefield, but a small fraction are caught up in it and are hugely accelerated, doubling in energy over a path of just 85 centimetres.

plasmas with modest energy gains[11]. Such beams will also need to be focused down to nanometre-spot sizes, perhaps using the transverse fields set up within a 'plasma lens'[12]. Efforts in these areas will receive a welcome boost in the wake of Blumenfeld and colleagues' work, which has brought electron plasma-wakefield accelerators to the energy frontier that was previously the undisputed territory of large-scale particle accelerators. ■

Robert Bingham is at the Rutherford Appleton Laboratory, Chilton, Didcot OX11 OQX, UK.
e-mail: r.bingham@rl.ac.uk

1. Blumenfeld, I. *et al. Nature* **445,** 741–744 (2007).
2. Mangles, S. *et al. Nature* **431,** 535–538 (2004).
3. Geddes, C. G. R. *et al. Nature* **431,** 538–541 (2004).
4. Faure, J. *et al. Nature* **431,** 541–544 (2004).
5. Hogan, M. J. *et al. Phys. Rev. Lett.* **95,** 054802 (2005).
6. Leemans, W. P. *et al. Nature Phys.* **2,** 696–699 (2006).
7. Faure, J. *et al. Nature* **444,** 737–739 (2006).
8. Dodd, E. S. *et al. Phys. Rev. Lett.* **88,** 125001 (2002).
9. Wang, S. *et al. Phys. Rev. Lett.* **88,** 135004 (2002).
10. Lee, S. *et al. Phys. Rev. ST Accel. Beams* **5,** 011001–011004 (2004).
11. Blue, B. E. *et al. Phys. Rev. Lett.* **90,** 214801 (2003).
12. Su, J. J., Katsouleas, T. & Dawson, J. M. *Phys. Rev. A.* **41,** 3321–3331 (1990).

VASCULAR BIOLOGY

Vessel guidance

Thomas Gridley

Embryos and tumours use the same signalling pathways to direct the formation of blood vessels. Discovery of a new role for the Notch pathway in that process presents a fresh option for cancer treatment.

Angiogenesis, the growth of new blood vessels from existing ones, is a highly active process in embryos. In adults, blood vessels in most organs are quiescent — except, notably, during the growth of solid tumours, when embryonic signalling pathways direct new blood vessels to grow around and into the tumour. Two players in this process are the vascular endothelial growth factor (VEGF) and Notch signalling pathways. Seven recent papers[1–7], including two on pages 776 and 781 of this issue[3,4], have yielded insights into Notch function during the formation of blood vessels in both embryos and tumours, and have revealed a new drug target for disrupting tumour angiogenesis.

VEGF is a secreted glycoprotein that is a potent inducer of angiogenesis both in embryos and in tumours[8,9]. The Notch pathway is an intercellular signalling system in which both the signalling (ligand) and receiving (receptor) molecules are anchored to the cell surface, thereby restricting signal transmission to cells that are physically adjacent. This pathway is frequently involved in regulation of cell differentiation, and previous work[10–12] had shown that the Notch ligand Delta-like 4 (Dll4) is essential for vascular development in mice. The new studies[3,4] identify a previously unknown role for Dll4/Notch signalling during vascular development, and clarify the mechanism responsible for the vascular defects that result from reduced Notch signalling.

The various groups analysed blood-vessel development in three different experimental systems: the zebrafish embryo[4,7], the retina of the mouse eye[2,3,5,6], and solid tumours growing in mice[1,2]. Developing zebrafish embryos are almost transparent, making them ideal for high-resolution imaging studies of blood-vessel development. The advantage of the mouse retina is that blood vessels develop mainly after birth in a highly reproducible spatial and temporal pattern; during these stages, the retinal vasculature is accessible both for observation and for experimental administration of drugs or other agents.

A finding common to all the studies[1–7] was that inhibition of Notch signalling led to increased sprouting and branching of blood vessels. The Notch pathway regulates sprouting and branching behaviours by influencing the formation of vascular 'tip cells' — specialized endothelial cells at the leading edge of vascular sprouts. The tip cells extend protrusions, called filopodia, that sense the local environment and guide growth of these sprouts along gradients of VEGF protein. In both the mouse retina and the zebrafish embryo, Dll4/Notch signalling regulated the formation of tip cells. Reduced Notch signalling led to increases in the number of tip cells, extension of filopodia and branching of vessels (Fig. 1). Conversely, pharmacological or genetic manipulations that blocked VEGF function reduced both Dll4 expression and blood-vessel sprouting[2,4–7], indicating that the suppression of tip-cell formation and angiogenic sprouting by Notch signalling occurs downstream of the VEGF signal.

Growth and metastasis of solid tumours require the recruitment of host blood vessels. Many solid tumours express the angiogenesis-promoting VEGF, and anti-VEGF therapies are effective in blocking growth of solid tumours in rats and mice[9,13]. Notch signalling is essential for angiogenesis in embryos[10–12], but does not seem to have a major role in maintaining established blood vessels in adults. So protein components of the Notch pathway, particularly if their expression and function are confined to the vascular system, may provide drug targets during tumour angiogenesis.

Two of the studies[1,2] have identified the Dll4 protein as just such a drug target. Systemic administration of either neutralizing antibodies against Dll4, or a recombinant form of the Dll4 protein that had been modified to block Dll4/Notch signalling[1], inhibited growth of several different solid tumours in mice. As with the findings in zebrafish embryos and mouse retinas, anti-Dll4 treatment increased the sprouting and branching of blood vessels, and led to a marked increase in blood-vessel density in tumours. But tests of the vascular network in these tumours revealed that the new vessels functioned inefficiently and were not connected functionally to the vascular network of the tumours, leading to an overall inhibition of tumour growth (Fig. 1).

Unfortunately for many patients with cancer, individual therapies can be ineffective against specific tumours, and tumours that respond initially to drugs can become resistant to them. In these studies[1,2], anti-Dll4 treatment inhibited tumour growth better when combined with anti-VEGF treatment than when given

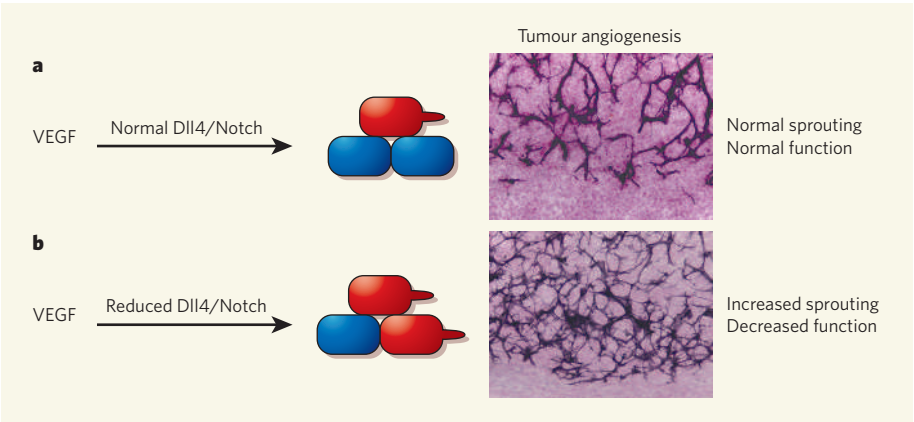


Figure 1 | The Notch pathway and angiogenesis. a, Vascular endothelial growth factor (VEGF) promotes Notch signalling, which normally suppresses formation of vascular tip cells (red). **b,** When Notch signalling is reduced (for example, by treatment with antibodies against Dll4), more tip cells form, reducing the numbers of non-tip cells (blue) and leading to increased blood-vessel sprouting and branching. But the new vessels function poorly, which in the case of tumour angiogenesis in mice stunts tumour growth[1,2]. (Photos from Fig. 2 of ref. 1.)

alone, and was effective even against tumours that did not respond to anti-VEGF therapies. Anti-Dll4 treatment may therefore provide a good option for alternative or combinatorial therapy for solid tumours that are resistant to anti-VEGF therapies.

A host of issues remains before anti-Dll4 treatments reach the clinic, however. For example, despite the effectiveness of treatment of rodent tumours with antibodies against VEGF, treatment with these agents in clinical trials of several types of cancer provided an overall survival benefit for patients only when they were combined with conventional chemotherapy[9,13]. Similar issues might arise as anti-Dll4 treatments progress into clinical trials. Nonetheless, with these new findings we now have another possible point of attack on cancer. ■

Thomas Gridley is at The Jackson Laboratory, Bar Harbor, Maine 04609, USA.
e-mail: tom.gridley@jax.org

1. Noguera-Troise, I. *et al. Nature* **444,** 1032–1037 (2006).
2. Ridgway, J. *et al. Nature* **444,** 1083–1087 (2006).
3. Hellström, M. *et al. Nature* **445,** 776–780 (2007).
4. Siekmann, A. F. & Lawson, N. D. *Nature* **445,** 781–784 (2007).
5. Lobov, I. B. *et al. Proc. Natl Acad. Sci. USA* (in the press).
6. Suchting, S. *et al. Proc. Natl Acad. Sci. USA* (in the press).
7. Leslie, J. D. *et al. Development* **134,** 839–844 (2007).
8. Coultas, L., Chawengsaksophak, K. & Rossant, J. *Nature* **438,** 937–945 (2005).
9. Jain, R. K., Duda, D. G., Clark, J. W. & Loeffler, J. S. *Nature Clin. Pract. Oncol.* **3,** 24–40 (2006).
10. Krebs, L. T. *et al. Genes Dev.* **18,** 2469–2473 (2004).
11. Duarte, A. *et al. Genes Dev.* **18,** 2474–2478 (2004).
12. Gale, N. W. *et al. Proc. Natl Acad. Sci. USA* **101,** 15949–15954 (2004).
13. Ferrara, N. & Kerbel, R. S. *Nature* **438,** 967–974 (2005).

QUANTUM MECHANICS

The truth about reality

Gregor Weihs

Hopes of keeping quantum mechanics 'real' have been dashed by new measurements of neutrons' quantum behaviour. Despite what our classical sensibilities require, the world is indeed fundamentally random.

Albert Einstein was convinced that "God does not play dice"; in other words, he could not accept quantum theory, with its inherent randomness, as a fundamental description of the world. Developing the theme in later work with Boris Podolsky and Nathan Rosen[1], he hinted that he believed in a more basic layer of truth underlying quantum mechanics. This was to be expressed in 'hidden variables' that reconciled the purely statistical validity of quantum measurements with the classical, deterministic world-view. Writing in *Physical Review Letters*, Yuji Hasegawa *et al.*[2] deal a further blow to these already beleaguered efforts to inject some realism into quantum physics.

Two theorems developed in the 1960s put severe constraints on attempts to complete quantum physics as Einstein intended. First, John Bell showed that theories of local hidden variables, which don't permit any remote influences, cannot explain certain quantum-physical observations[3]. Second, Simon Kochen and Ernst Specker independently proved that more general, so-called non-contextual hidden variables (of which more later) are also untenable[4]. Many experiments have since used Bell's theorem to invalidate local hidden variables. Much less work has been done on the Kochen–Specker theorem, especially for particles other than photons.

Neutrons are convenient guinea-pigs for the kind of delicate experiments needed to investigate these aspects of quantum physics: above all, they have no charge, which often makes it easier to observe the effects involved. Chief among these are the effects of spin, the property of a particle that makes it try to line up in a magnetic field. Quantum physics tells us that spin will be quantized: if you measure it along a chosen direction, it will point either in that direction or in the opposite one, but never in between. In addition, if you have measured spin in the vertical (z) direction and found it to be up, a subsequent measurement in the horizontal direction (x) will randomly yield a right or left value. Similarly, if you measure x first, the z value will be random.

It is this kind of randomness that was not to the taste of physicists such as Einstein. Hidden variables would allow the result of the z and the x measurements to be determined simultaneously. In stark contrast, quantum physics nonchalantly declares that the z and x components of the spin cannot be quantified at the same time. It is, in fact, a matter of principle that the two measurements require different experiments.

Other characteristics, such as a neutron's position, can be measured at the same time as spin. The procedures for measuring position and spin don't interfere with each other, and the measurements are said to be compatible. It would therefore be natural to want any hidden variables that are introduced to explain the spin-component measurements also to explain the results of simultaneous spin and position measurements: the value governed by the spin's hidden variable shouldn't depend on a specific position measurement, and vice versa. This is the condition known as non-contextuality: that the result yielded by any hidden variable should not differ according to any compatible measurement being performed at the same time.

But it seems that, as always, quantum theory wants to have the last word: it stubbornly refuses to admit hidden variables even under such seemingly innocent conditions. It turns out that neutrons can be prepared in such a way that spin and position measurements, although nominally still independent, are so strongly correlated that non-contextual hidden variables cannot explain them. Unsurprisingly for connoisseurs of quantum weirdness, entanglement — the mysterious holism in which the state of one quantum object is tied to the state of a second, separate object — is the key to this trick.

The fact that non-contextual hidden variables cannot explain the spin–position correlations of entangled neutrons is a variant of the Kochen–Specker theorem. It was long thought to be untestable because, in its original form, it required infinitely precise measurements. With a statistical treatment, however, an inequality could be derived[5–7] that restricted the

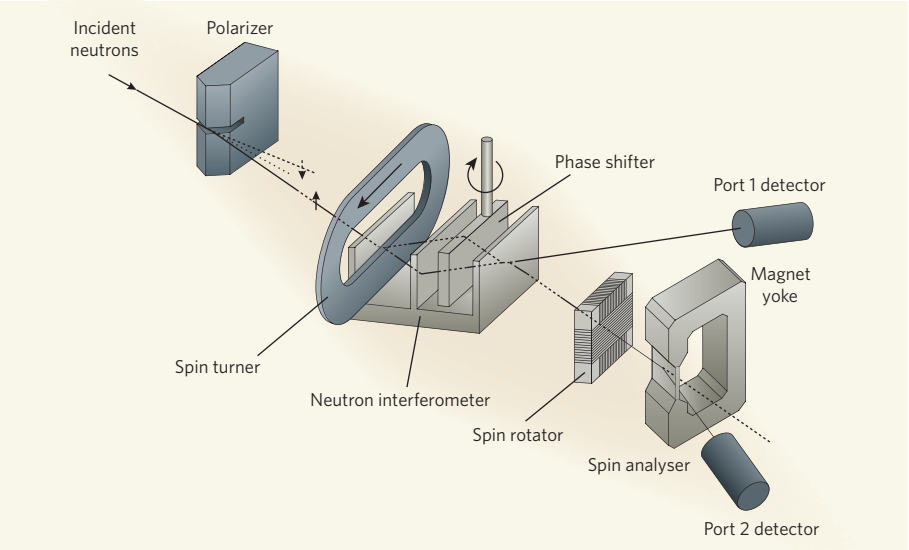


Figure 1 | Neutron interferometer. Hasegawa and colleagues' latest experiments[2] on quantum reality reuse an earlier experimental set-up of theirs[8]. Neutrons are spin-polarized before they enter the interferometer, where the neutron beam is split into two possible paths before recombining and exiting through one of two ports. An entanglement between path and spin is established by changing the spin of the neutrons on one of the paths. The phase shifter determines the type of port measurement by exploiting neutron interference at the point where the two beams recombine. A spin rotator after the interferometer selects the direction in which spin is measured for one of the beams. The correlations between spin and exit port measurements can be determined from the numbers of neutrons detected per unit time for each setting. (Figure after ref. 8.)

predictions for measurement results made by any non-contextual hidden variable model.

With this theory at hand, Hasegawa *et al.*[2] passed a beam of neutrons through a specially designed interferometer that splits the incoming neutrons into two beams with opposing spin states (Fig. 1). The path that a particular neutron takes thus becomes entangled with its spin. The paths are recombined after the neutron passes a phase shifter (inducing a variable delay), and the neutron leaves the interferometer through one of two exit ports, depending on its phase shift. The end result is that you can measure a neutron's position — the port it exits from — as well as its spin along one of two orthogonal directions.

The exit port measurements for two different phase shifts follow the same rules as z and x spin measurements. You cannot measure both simultaneously, and if you measure one after the other, the second measurement will randomly reveal the neutron in one of the two ports. By analogy, we could call them the port z and x measurements.

Hasegawa and colleagues performed z and x measurements of both the spins and the exit ports of the neutrons as they emerged from the interferometer. To no great surprise, they demonstrated convincingly that the correlations between the measurements are stronger than anything allowed by the theoretical inequality[5,6] for models involving non-contextual hidden variables. In essence, the spin measurements fix the spin values of the hidden variables, and the port measurements fix the port values. That leaves no more room for manoeuvre. But for the hidden variables to be compatible with the result of the joint measurement of spin and port, more latitude is just what is required.

Even though experiments of this type are very difficult — Hasegawa and colleagues' neutron beam is about as dim as the light of a candle seen 16 kilometres away — the data are amazingly clean and leave nothing to interpretation. Thus, unless one allows the existence of contextual hidden variables with very strange mutual influences, one has to abandon them — and, by extension, 'realism' in quantum physics — altogether. We knew this for photons already, but the corroboration in a different system should help to convince doubting Thomases, as well as assure the rest of us.

At a time when quantum information processing is becoming an established field in physics and computer science, it is important to experiment on the foundations of the underlying theory. This might seem strange, given that the past 100 years have shown that quantum theory is very good at predicting the results of experiments. But if you agree with Einstein, then it doesn't matter how practical the theory is: it can still just not real be enough. ■

Gregor Weihs is at the Institute for Quantum Computing, and Department of Physics and Astronomy, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada.
e-mail: weihs@iqc.ca

1. Einstein, A., Podolsky, B. & Rosen, N. *Phys. Rev.* **47,** 777–780 (1935).
2. Hasegawa, Y. *et al. Phys. Rev. Lett.* **97,** 230401 (2006).
3. Bell, J. *Physics* **1,** 195–200 (1964).
4. Kochen, S. & Specker, E. *J. Math. Mech.* **17,** 59–87 (1967).
5. Cabello, A. & García-Alcaine, G. *Phys. Rev. Lett.* **80,** 1797–1799 (1998).
6. Simon, C., Weihs, G. & Zeilinger, A. *Phys. Rev. Lett.* **84,** 2993–2996 (2000).
7. Simon, C., Brukner, D. & Zeilinger, A. *Phys. Rev. Lett.* **86,** 4427–4430 (2001).
8. Hasegawa, Y., Loidl, R., Badurek, G., Baron, M. & Rauch, H. *Nature* **425,** 45–48 (2003).

DEVELOPMENTAL BIOLOGY

Cell fate in the mammary gland

Qiang Tong and Gökhan S. Hotamisligil

Most breast cancers have their origin in the luminal epithelial cells of the mammary gland. Defining how a master regulator controls the development of this cell lineage could provide important hints about why this should be.

The developmental biology of the mammary gland is a fascinating process — from its origins in the embryo to the many dynamic changes in later stages of life. But it is thought to be perverted in breast cancer. The gland consists of a tree-shaped network of ducts, with the outer layer of the tubes consisting of myoepithelial cells, and epithelial cells lining the inner lumen. During pregnancy, some of the luminal epithelial cells develop into lobuloalveolar cells, which, given the right hormonal cues, secrete milk. The overwhelming majority of breast cancers arise from these luminal cells.

In two new papers, Kouros-Mehr *et al.*[1] and Asselin-Labat *et al.*[2] show that a gene-regulatory factor called GATA-3 guides immature breast cells on their developmental path to form luminal epithelial cells. Moreover, they have uncovered a link between GATA-3 and two additional gene-regulatory factors, FoxA1 and oestrogen receptor alpha (ERα). Their results indicate that this pathway may be involved in mammary tumours.

The GATA-3 protein is an old acquaintance of developmental biologists, as it regulates the lineage determination and differentiation of many cell types. These include neurons of the sympathetic nervous system, immune cells called T helper 2 (T_H2) cells, the inner root sheath of hair follicles, fat cells, the nephritic ducts of the kidney, and the ear cochleae[3,4]. Because GATA-3 is such a central player in development, fetal mice cannot survive to term without it[5], and in humans a mutation in one of the copies of the GATA-3 gene leads to many

abnormalities, including problems with the thyroid gland, the kidneys and hearing[6].

Normally, GATA-3 is expressed at high levels, together with ERα, in the luminal epithelial cells of the mammary gland. Interestingly, in breast cancer, higher GATA-3 expression in these cells predicts a better prognosis[7]. Kouros-Mehr *et al.*[1] and Asselin-Labat *et al.*[2] found that GATA-3 expression occurs early in embryonic mammary development (at embryonic day 12.5), but only in the luminal epithelia (ductal and lobuloalveolar cells), not in the myoepithelium.

To find out how GATA-3 functions in mammary-gland development, the authors created a complex set of genetically engineered mice. They used tissue-specific gene-deletion technology to introduce mutations into the GATA-3 gene in subsets of cells and/or at certain developmental stages. The experiments revealed that GATA-3 expression is required for normal embryonic mammary development. In the postnatal mammary gland, GATA-3 is essential for the formation of the ductal tree and for differentiation of the epithelial cells.

Deleting the GATA-3 gene in lobuloalveolar 'units' in pregnant mice disrupted the development of the units and milk production. Importantly, these defects occurred mainly because the lobuloalveolar cells did not develop properly, rather than because of cell death or abnormalities in cell proliferation. This loss of GATA-3 expression also resulted in diminished expression of several of the protein markers that characterize the luminal epithelia, including oestrogen receptors[1,2]. Notably, the effects of GATA-3 deficiency on the mammary gland resemble those of ERα deficiency[8], so these two proteins may well belong to the same molecular or functional pathway.

Asselin-Labat *et al.*[2] also defined a mammary-cell differentiation scheme in terms of the cell-surface proteins expressed by the cells at various stages — from primitive and unspecialized stem cells, through the intermediate, partially committed progenitor cells, to mature luminal epithelial cells. Having defined the different developmental stages, the authors showed that levels of GATA-3 and ERα expression are lowest in the least-developed cells, and gradually increase during differentiation, being highest in mature luminal cells. Moreover, deleting GATA-3 from progenitor cells blocked any differentiation into luminal cells, and the forced expression of GATA-3 in mammary stem-cell-enriched populations promoted differentiation into luminal cells[2]. So it seems that GATA-3 is not merely required to maintain the luminal epithelial lineage, but actively determines the fate of these cells.

With the function of GATA-3 in luminal epithelial differentiation firmly established, it became obvious that this protein might have a role in breast tumorigenesis. The luminal A subtype of breast tumour has a relatively favourable prognostic outcome — and the highest GATA-3 and ERα expression levels of all breast tumours[7]. In luminal B tumours, GATA-3 and ERα levels are markedly reduced, whereas basal-like tumours, which have the worst prognosis, have the lowest expression of GATA-3 and ERα. In fact, GATA-3 seems to be an even better prognostic indicator than the oestrogen receptor, a current prognostic marker[7]. Furthermore, mutations in the GATA-3 gene have been identified in various breast tumours, suggesting that the protein might act as a tumour suppressor.

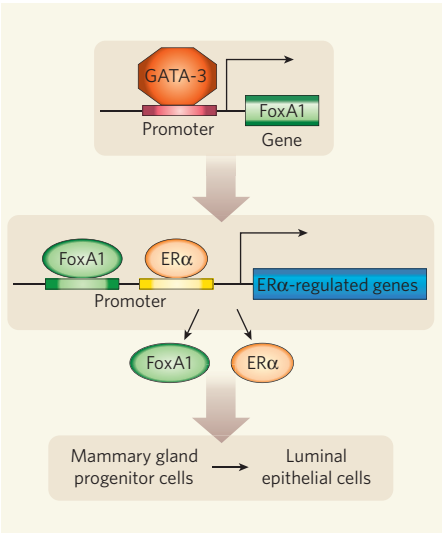


Figure 1 | Role of GATA-3 in mammary-gland development. Kouros-Mehr *et al.*[1] and Asselin-Labat *et al.*[2] establish that GATA-3 specifies the fate of progenitor cells in the mammary gland, directing them to differentiate into luminal epithelial cells. Kouros-Mehr *et al.*[1] elucidate a pathway from GATA-3 to oestrogen receptor alpha (ERα), in which GATA-3 binds to the regulatory region (promoter) of the gene encoding FoxA1 and possibly activates its expression. FoxA1 might then interact with ERα and bind to the ERα gene promoter to turn on the transcription of ERα target genes, which include FoxA1 and ERα. Perturbation of this pathway as a result of decreased GATA-3 expression may contribute to breast tumours.

Low levels of GATA-3 expression in certain tumours might indicate that the cells have lost some properties of differentiation, or that the tumours originated from undifferentiated stem cells. Alternatively, as Kouros-Mehr *et al.*[1] suggest, a lack of GATA-3 might actually encourage cell proliferation. In further mechanistic studies, they show that GATA-3 binds to the regulatory region of the gene encoding FoxA1, another gene-regulatory factor, and that expression of FoxA1 is highly correlated with the level of GATA-3 in the mammary gland. FoxA1 might interact physically with ERα (ref. 9). Moreover, FoxA1-binding sites on chromosomes often lie next to the gene-regulatory sequences targeted by ERα, and the FoxA1–ERα interaction enhances the ability of ERα to gain access to these sites[10]. So, a reduction of FoxA1 might hinder the action of the oestrogen receptor, which is required for ductal- and lobuloalveolar-cell development and drives cell proliferation in breast tumours (Fig. 1). The story promises to be far more complicated, however, as high levels of FoxA1 also seem to have the ability to inactivate the oestrogen receptor[11]. The intricacies of these pathways, and potential differences of regulation between normal and tumour cells, need further clarification before we know whether GATA-3 could be used to manage breast cancer.

There are similarities in the genetic control of the differentiation programmes of T helper cells, hair follicles, cells of connective tissue and mammary epithelium. In all these systems, GATA-3 is involved in directing the development of one cell lineage from progenitor cells, usually at the cost of another lineage[3,4]. For instance, in T-helper-cell development, T_H1 and T_H2 cells are derived from a common type of progenitor cell. GATA-3 drives T_H2-cell formation and inhibits the differentiation of T_H1 cells[3]. In adipose tissue, GATA-3 inhibits the fat-generating factors PPARγ and C/EBPα, and it directs progenitor cells towards the formation of bone and possibly other tissues[12,13]. In the mammary gland, whereas GATA-3 seems to specify the luminal epithelial-cell lineage, it is unclear which gene regulates the differentiation of myoepithelium. It seems that loss of GATA3 does not strongly influence this process.

Several other lines of enquiry present themselves. For example, so-called friend-of-GATA factors (FOG1 or FOG2) usually function in concert with GATA gene-transcription factors[14]. So, do these factors also have a role in mammary-gland development, plasticity or tumorigenesis? And do other developmental pathways, such as the hedgehog and insulin/IGF signalling pathways, which regulate GATA factors in other lineage specification schemes[14], interact with GATA-3 in mammary development? Finally, given the role of GATA-3 in adipocyte differentiation[12], does GATA-3 activity in the stroma and fat pad also contribute to mammary development? Whatever the answers, GATA-3 is now established as a master regulator of the mammary gland. ■

Qiang Tong is in the USDA/ARS Children's Nutrition Research Center, Baylor College of Medicine, Houston, Texas 77030, USA.
Gökhan S. Hotamisligil is in the Department of Genetics and Complex Diseases, Harvard School of Public Health, Boston, Massachusetts 02115, USA.
e-mails: qtong@bcm.tmc.edu; ghotamis@hsph.harvard.edu

1. Kouros-Mehr, H., Slorach, E. M., Sternlicht, M. D. & Werb, Z. *Cell* **127,** 1041–1055 (2006).
2. Asselin-Labat, M. L. *et al. Nature Cell Biol.* **9,** 201–209 (2007).
3. Pai, S. Y., Truitt, M. L. & Ho, I. C. *Proc. Natl Acad. Sci. USA* **101,** 1993–1998 (2004).
4. Kaufman, C. K. *et al. Genes Dev.* **17,** 2108–2122 (2003).
5. Pandolfi, P. P. *et al. Nature Genet.* **11,** 40–44 (1995).
6. Van Esch, H. *et al. Nature* **406,** 419–422 (2000).
7. Mehra, R. *et al. Cancer Res.* **65,** 11259–11264 (2005).
8. Mallepell, S., Krust, A., Chambon, P. & Brisken, C. *Proc. Natl Acad. Sci. USA* **103,** 2196–2201 (2006).

9. Schuur, E. R. *et al. J. Biol. Chem.* **276,** 33554–33560 (2001).
10. Carroll, J. S. *et al. Cell* **122,** 33–43 (2005).
11. Wolf, I. *et al. Int. J. Cancer* **120,** 1013–1022 (2007).
12. Tong, Q. *et al. Science* **290,** 134–138 (2000).
13. Suh, J. M. *et al. Cell Metab.* **3,** 25–34 (2006).
14. Cantor, A. B. & Orkin, S. H. *Semin. Cell Dev. Biol.* **16,** 117–128 (2005).

NANOFLUIDICS

Silicon for the perfect membrane

Albert van den Berg and Matthias Wessling

Newly developed ultrathin silicon membranes can filter and separate molecules much more effectively than conventional polymer membranes. Many applications, of economic and medical significance, stand to benefit.

On page 749 of this issue, Striemer *et al.*[1] describe a method for preparing ultrathin nanoporous membranes made from silicon. Nanoporous membranes are already widely used in medicine, for instance for the filtration and separation of blood proteins in an artificial kidney (haemodialysis) — a rapidly growing world market currently worth more than US$1 billion annually. They can also function as a mechanical support for desalination membranes used to purify sea water for irrigation and human consumption. Given that the membrane technology is seemingly so mature, why should we bother searching for new methods and different starting materials?

At present, all technologically relevant nanoporous membranes are prepared by initiating the precipitation of a polymer from solution. This is achieved through the addition of a non-solvent (often water), or by rapid cooling. The solution precipitates into micrometre- and nanometre-sized domains rich in polymer that form a filter structure. Between these polymer domains, polymer-free areas form the pore system. A diverse spectrum of morphologies and geometries can thus be produced from a variety of starting materials[2].

These nanoporous membranes have a thin skin, typically less than 500 nanometres thick, made up of small bumps, or nodules, with a radius of a few to 50 nm. The voids between the nodules determine the pore size; the pores are 1–50 nm across, and thus the porosity of the membrane as a whole is low. A much thicker layer, with a larger pore size and porosity, provides mechanical support for the nodular skin. Although the pore size of the membrane skin can be adjusted by the choice of starting material and processing route, other morphological parameters, such as its thickness, porosity and pore-size distribution, are surprisingly insensitive to such choices[3].

Nanoporous membranes prepared according to these methods suffer from a typical trade-off: the flux through them can be enhanced by increasing the pore diameter, but at the cost of less effective molecular discrimination. Optimizing flux and selectivity simultaneously requires a fundamentally new approach, which Striemer and colleagues[1] offer.

Not only do the authors' porous nanocrystalline silicon (pncSi) membranes combine small membrane thickness and pore sizes (Fig. 1), but they are also robust, their pore size can be controlled, and they are simple to produce. Earlier attempts to make ultrathin nanoporous membranes used either sophisticated nanolithography or were based on colloidal templates[4,5]. The first method is expensive; and although the second makes elegant use of self-organization principles, very small, controlled pore sizes are difficult to achieve.

Striemer and colleagues' nanopores self-form from a deposited layer of amorphous silicon through rapid thermal annealing[6]. The pore sizes can be controlled between 5 and 25 nm (the range of interest for protein separation) by the choice of annealing temperature. Although the pore-size distribution is not extremely narrow, it has no tail to larger pore sizes. The absence of such a tail is a prerequisite for molecular specificity — and still a challenge for state-of-the-art polymer-based membranes.

The authors find that two important proteins, immunoglobulin-γ and bovine serum albumin (BSA), with hydrodynamic diameters of 14 and 6.8 nm, and molecular weights that similarly differ by a factor of a little more than two, can be separated using their pncSi membrane. For efficient separation using conventional ultrafiltration membranes, a molecular-weight ratio of more than ten is needed. The flux through the pncSi membranes is more than ten times faster than that through conventional membranes with similar selectivity properties. Moreover, Striemer *et al.*[1] find that by changing the surface charge of their membrane through chemical modification, they can separate proteins that are similar in size, but bear a different charge[7].

Perhaps the most promising advantage of the method presented by Striemer *et al.* is that it can be easily integrated into 'labs-on-a-chip' — microfluidics systems that are currently enjoying rapidly growing attention owing to their potential for medical diagnostics, drug discovery and chemical synthesis[8–10]. It is that promise of integration with other nanofluidic separation and analysis techniques for biochemical and biomedical applications that, together with the inherent advantages of the silicon-based system, make this such an important step forward. We look forward to further improvements and proposals for additional uses for the technique. ■

Albert van den Berg and Matthias Wessling are at the MESA+ Institute for Nanotechnology and the Faculty of Electrical Engineering, Computer Science and Mathematics, and the Faculty of Science and Technology, University of Twente, PO Box 217, Enschede, the Netherlands.
e-mail: a.vandenberg@utwente.nl

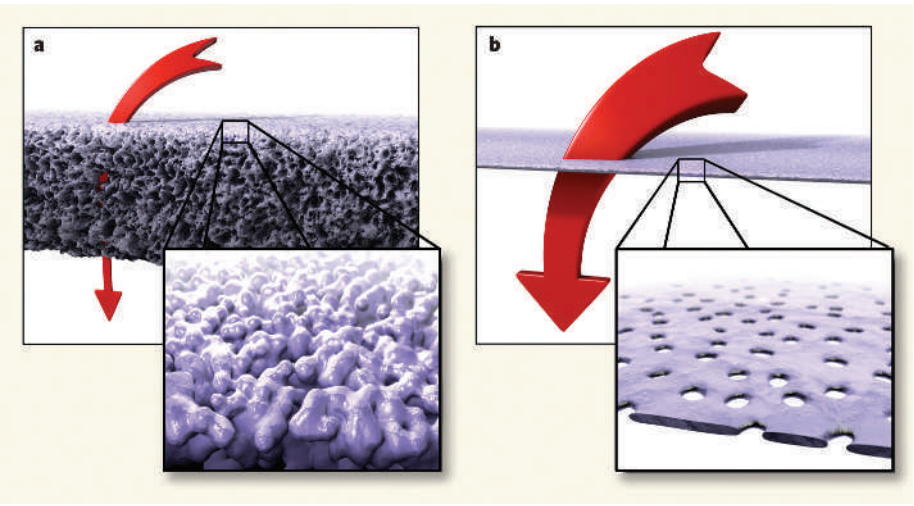


Figure 1 | Barrier to progress. a, The nanoscale nodules that make up the conventional ultrafiltration membrane form a significant restriction to flow. **b,** The ultrathin porous nanocrystalline silicon (pncSi) membranes developed by Striemer *et al.*[1] allow efficient protein separation without restricting the flow as much.

1. Striemer, C. C., Gaborski, T. R., McGrath, J. L. & Fauchet, P. M. *Nature* **445,** 749–753 (2007).
2. Vogelaar, L. *et al. Small* **1,** 645–655 (2005).
3. Mehta, A. & Zydney, A. L. *J. Membr. Sci.* **249,** 245–249 (2005).
4. Tong, H. D. *et al. Nano Lett.* **4,** 283–287 (2004).
5. Yan, F. & Goedel, W. A. *Adv. Mater.* **16,** 911–915 (2004).
6. Grom, G. F. *et al. Nature* **407,** 358–361 (2000).
7. Eijkel, J. C. T. & van den Berg, A. *Lab Chip* **6,** 19–23 (2006).
8. Whitesides, G. M. *Nature* **442,** 368–373 (2006).
9. van den Berg, A. & Bergveld, P. *Lab Chip* **6,** 1266–1273 (2006).
10. de Jong, J. *et al. Lab Chip* **6,** 1125–1139 (2006).

STRUCTURAL BIOLOGY

Analysis of 'downhill' protein folding

Arising from: M. Sadqi, D. Fushman & V. Muñoz *Nature* **442**, 317–321 (2006)

There is controversy as to whether homologues from the peripheral subunit binding domain family of small proteins fold 'downhill' (that is, non-cooperatively, in the absence of free-energy barriers between conformations) and whether they modulate their size for biological function. Sadqi *et al.*[1] claim that Naf-BBL — a naphthylalanine-labelled, truncated version of this domain — folds in this way, on the grounds that they recorded a wide spread of melting temperatures of individual atoms measured by proton nuclear magnetic resonance (NMR) during their thermal denaturation. But their data are not of adequate quality to distinguish, within experimental error, between downhill folding and folding with a cooperative transition. Accordingly, their results offer no compelling evidence that Naf-BBL folds downhill, particularly as non-truncated, unmodified peripheral subunit binding domains seem to fold cooperatively[2,3].

Central to the case for non-cooperative folding is a wide dispersion of melting temperatures for individual atoms. The representative NMR titrations supporting such a dispersion are presented as normalized curves by Sadqi *et al.*[1] (Fig. 1a), without an error analysis and with no indication of the amplitudes and baselines that are essential components for normalization. From their Supplementary Information[1], we find that the proton chemical shifts were measured over too narrow a temperature range to capture the crucial baselines and that the amplitudes of many curves were far too small, yielding large uncertainties in curve fitting[4].

The authors claim that they resolved the melting temperatures of 158 sets of chemical shifts and fitted them to two- or three-state models, using "physically reasonable" baselines as constraints[1]. However, BBL has a structured denatured state[3]. The gradual melting of the residual structure, together with temperature effects on the native structure (such as the fraying of the amino and carboxyl termini — for example, lysine residue 40 and its neighbours in Naf-BBL), contribute to the chemical-shift changes and may give rise to 'unreasonable' baselines and anomalous titrations. Further, changes in ionization states with temperature may cause electrostatic and local structural changes that affect neighbouring chemical shifts: for example, Naf-BBL contains two histidine residues. Sadqi *et al.*[1] attempt to select midpoints by using the derivatives of the plots, but this approach is problematic, given the paucity of data and the sensitivity to experimental error within the broad transitions, especially for those of small amplitude (Fig. 1).

We attempted to fit all those data to the standard two-state equation[5] without bias by using a standard Marquardt algorithm with no constraints imposed on baselines. We found only about 70 curves that had a fitting standard error of less than 5 K, which were grouped around a mean of ~304 K and an enthalpy of denaturation of about 20 kcal mol^{-1} at the melting temperature. But the data are generally too poor to be reliable. We then fitted each of the 158 curves individually to a single melting temperature of 304 K and an enthalpy of denaturation of 20 kcal mol^{-1}, values determined from the free-fitted data of highest quality. The fits are acceptable (Fig. 2), except for those sensitive to ionization and fraying of termini.

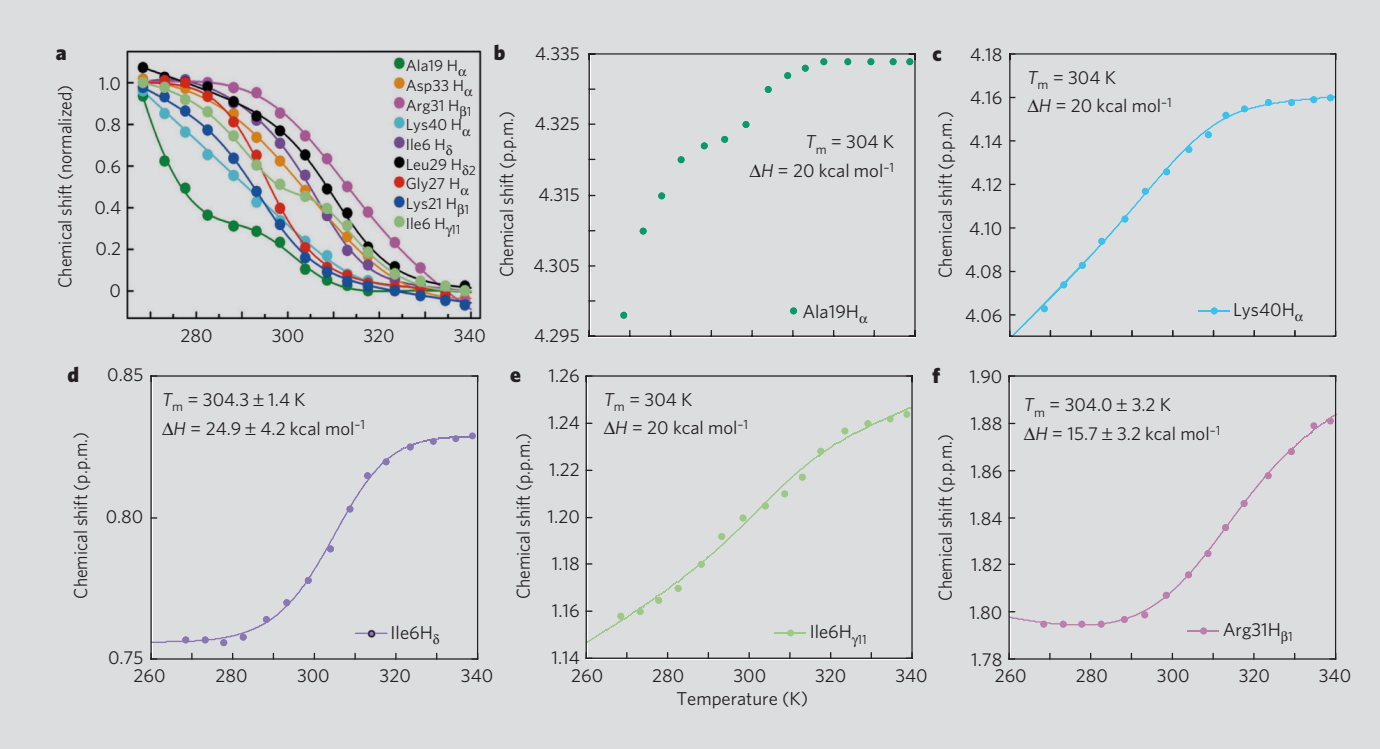


Figure 1 | Normalized and raw data for chemical shifts. a, The normalized data of Sadqi *et al.*[1] show a wide spread of melting temperatures (T_ms) for unfolding Naf-BBL. **b–f,** The unprocessed chemical shifts, however, reveal that: Ala19H_α (**b**) has too small an amplitude to be analysed, and is close to the ionizable residue His37 in the highest-resolution structure available (PDB code, 1w4h); Lys40H_α (**c**), at the dynamic carboxyl terminus, has merged baseline and transition and fits well to a T_m of 304 K; Ile6H_δ (**d**) free-fits well to 304 K; the small deviations in titration of Ile6$H_{\gamma 11}$ (**e**) are within experimental error of two-state behaviour; and Arg31$H_{\beta 1}$ (**f**) free-fits to a T_m of 304 K, rather than the ~313 K presented in **a**. Proton ('H') positions in an amino acid (three-letter notation) at a particular residue number are indicated by their subscript.

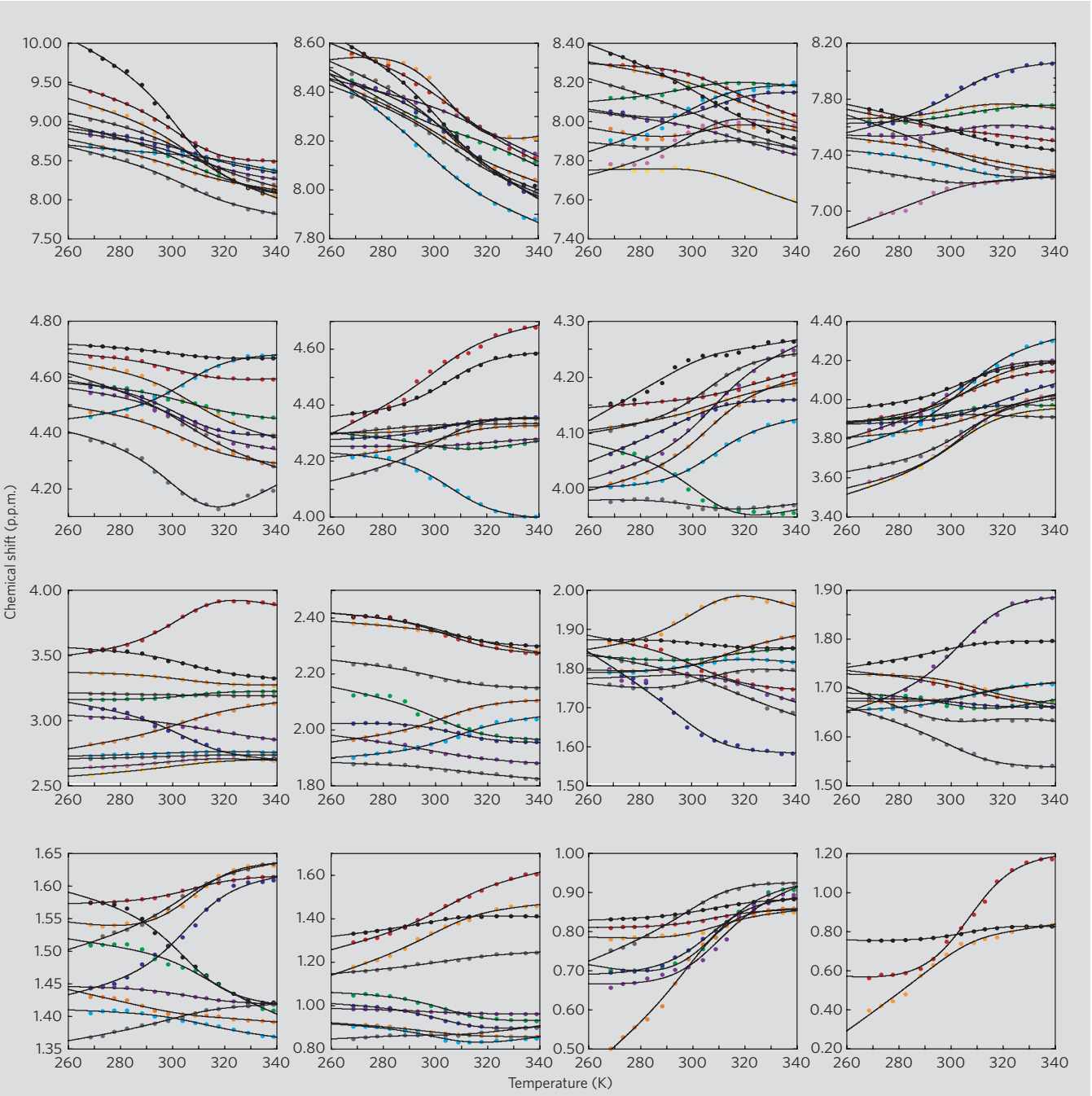


Figure 2 | Whole data set fits. The whole data set is fitted to a conventional cooperative transition of a melting temperature of 304 K, an enthalpy of denaturation at the melting temperature of 20 kcal mol^{-1} and a change in heat capacity of 350 kcal mol^{-1} K^{-1}. Data are grouped according to chemical shift.

The absence of well defined baselines allows substantial latitude in fitting to different models. By choosing arbitrary baselines and normalizing data, the resulting curves can be fitted to any desired model. Our two-state fit was merely to show that the data can be fitted readily to models other than downhill folding. A three-state model would accommodate the entire set even better. Indeed, we have been punctilious in describing the thermal unfolding transition of BBL as barrier-limited, rather than two-state[3].

The downhill-folding hypothesis changes the view of protein folding and calls for evidence that is beyond statistical dispute. Why do Sadqi *et al.*[1] study an unnaturally truncated, destabilized and chemically modified version of a protein to test protein-folding mechanisms and their claim of a biological role for peripheral subunit binding domains as molecular rheostats? They study Naf-BBL under strongly destabilizing conditions at a pH of 5.3, at which the melting temperature is some 20 K lower than at a neutral pH, thereby significantly truncating the crucial native baselines. By contrast, the more stable, full-length, unmodified domains have sharper transitions with long, native-state baselines that provide more accurate data and fold cooperatively[2,3].

Neil Ferguson, Timothy D. Sharpe, Christopher M. Johnson, Pamela J. Schartau, Alan R. Fersht

Medical Research Council Centre for Protein Engineering, Medical Research Council Centre, Cambridge CB2 2QH, UK
e-mail: arf25@cam.ac.uk

1. Sadqi, M., Fushman, D. & Muñoz, V. *Nature* **442,** 317–321 (2006).
2. Ferguson, N., Schartau, P. J., Sharpe, T. D., Sato, S. & Fersht, A. R. *J. Mol. Biol.* **344,** 295–301 (2004).
3. Ferguson, N. *et al. J. Mol. Biol.* **353,** 427–446 (2005).
4. Religa, T. L., Markson, J. S., Mayor, U., Freund, S. M. V. & Fersht, A. R. *Nature* **437,** 1053–1056 (2005).
5. Nicholson, E. M. & Scholtz, J. M. *Biochemistry* **35,** 11369–11378 (1996).

Competing financial interests: declared none.
doi: 10.1038/nature05643

STRUCTURAL BIOLOGY

Analysis of protein-folding cooperativity

Arising from: M. Sadqi, D. Fushman & V. Muñoz *Nature* **442**, 317–321 (2006)

The folding of small proteins has been assumed to be an all-or-none process that involves high cooperativity within the structure and substantial kinetic-energy barriers. Sadqi *et al.*[1] claim that the small re-engineered protein Naf-BBL unfolds without significant cooperativity or kinetic hindrance, a conclusion that is based on calculation of a broad distribution of midpoint thermal-transition temperatures measured by the nuclear magnetic resonance (NMR) chemical shifts of 158 protons. We find that all of the unprocessed melting curves can be fitted to the same two-state global unfolding when uncertainties in the experimental data are taken into account. We conclude that the authors' melting data for Naf-BBL remain consistent with the all-or-none process.

Figure 1a shows the normalized melting curves of typical protons, which were used by Sadqi *et al.* to illustrate the spread in melting temperatures; Fig. 1b shows the global two-state unfolding curve, which has a melting temperature of 304 K, a change in heat capacity of 0.4 kcal mol^{-1} K^{-1}, and a melting enthalpy of 20.0 kcal mol^{-1} (refs 2,3), demonstrating a lack of well defined baselines. In their analysis, Sadqi *et al.* use chemical shifts of random coils to mimic the post-transition baselines, but their melting temperatures, melting enthalpy and pre-transition baselines were determined by fitting the experimental data to a two- or three-state model. The melting temperatures and baselines obtained from the data fitting were subsequently used to calculate an index for evaluating unfolding cooperativity and a matrix of thermodynamic coupling for a high-resolution description of the unfolding process.

This analysis by Sadqi *et al.*[1] is problematic, because the lack of well defined baselines unavoidably causes important uncertainties in the melting temperatures and pre-transition baselines determined from the data fitting. In addition, the post-transition baselines are likely to be not properly represented by the chemical shifts of random coils because the unfolded states of proteins commonly deviate from those of random coils[4]. Sadqi *et al.* also use derivatives of melting curves to determine melting temperatures, which are independent of baselines. However, this method has limitations and could lead to more than one melting temperature for a two-state unfolding curve with small experimental errors.

Figure 1c–g shows the two-state fitting for the unprocessed data of typical protons with the same thermodynamic parameters as for global unfolding, and properly chosen baselines. The proton on the nitrogen atom of the amino acid threonine at residue 23 ($Thr23H_N$; Fig. 1c) shows the largest change in chemical shift during melting and is a good fit to the two-state model; $Ile6H_{\gamma 11}$ (Fig. 1d) fits the two-state model with small deviations. Sadqi *et al.* consider that $Ile6H_{\gamma 11}$ unfolds in a three-state manner, but they provide no error analysis or objective criteria for its classification. $Lys40H_\alpha$ (Fig. 1e) and $Arg31H_{\beta 1}$ (Fig. 1f) fit nicely to the two-state model. In the case of $Lys40H_\alpha$, the pre-transition baseline is closely merged to the transition region, which would make it difficult to obtain a reliable melting temperature and pre-transition baseline in the data fitting by Sadqi *et al.*

Both $Ala19H_\alpha$ (Fig. 1g) and $Arg28H_N$ (Fig. 1h) seem to deviate substantially from the global melting curve in the normalized

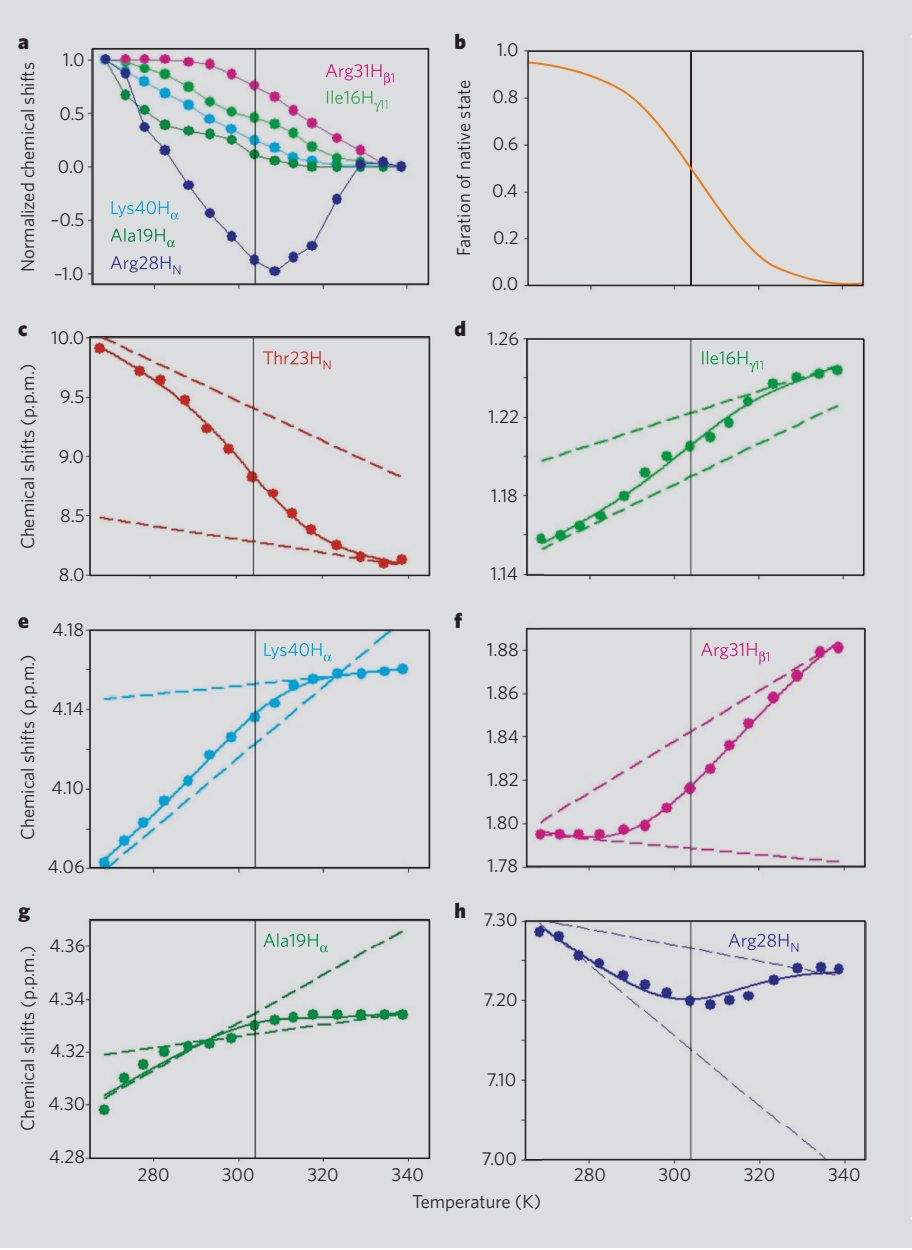


Figure 1 | Analysis of the melting curves of Naf-BBL. a, Normalized melting curves of typical protons. **b,** Global melting curve of Naf-BBL. **c–h,** Melting curves of typical protons. Solid lines are the two-state fitting curves and dashed lines are the chosen baselines; vertical lines are set at 304 K. Proton ('H') positions in an amino acid (three-letter notation) at a particular residue number are indicated by their subscript.

plot (Fig. 1a, b). However, they can actually be fitted to it, with small deviations, because they have very small melting amplitudes. The small deviations could arise from several factors, including experimental error, changes in ionization state with temperature, and residual structures in the unfolded state. Although the experimental data that do not fit a two-state model should not be ignored, this re-examination shows that the origins of the deviations should also be considered before attributing them to unfolding events.

The good fit of the 158 melting curves to the global two-state unfolding in the absence of well defined baselines does not prove that Naf-BBL unfolds in a two-state manner. However, it does show that highly cooperative unfolding cannot be excluded as an alternative model for the melting data. Accordingly, the non-cooperative downhill folding provides neither a reliable high-resolution description of, nor a necessary model for, the unfolding of Naf-BBL. An ultimate determination of the unfolding cooperativity of Naf-BBL requires the measurement of both pre- and post-transition baselines.

Zheng Zhou, Yawen Bai

Laboratory of Biochemistry and Molecular Biology, Center for Cancer Research, NCI, NIH, Building 37, Room 6114E, Bethesda, Maryland 20892, USA
e-mail: yawen@helix.nih.gov

1. Sadqi, M., Fushman, D. & Muñoz, V. *Nature* **442,** 317–321 (2006).
2. Ferguson, N., Schartau, P. J., Sharpe, T. D., Sato, S. & Fersht, A. R. *J. Mol. Biol.* **344,** 295–301 (2004).
3. Naganathan, A. N., Perez-Jimenez, R., Sanchez-Ruiz, J. M. & Muñoz, V. *Biochemistry* **44,** 7435–7449 (2005).
4. Dyson, H. J. & Wright, P. E. *Adv. Protein Chem.* **62,** 311–340 (2002).

Competing financial interests: declared none.
doi:10.1038/nature05644

STRUCTURAL BIOLOGY

Sadqi et al. *reply*

Replying to: N. Ferguson, T. D. Sharpe, C. M. Johnson, P. J. Schartau & A. R. Fersht *Nature* 445, doi: 10.1038/nature05643 (2007); Z. Zhou & Y. Bai *Nature* 445, doi: 10.1038/nature05644 (2007)

Ferguson *et al.*[1] and Zhou and Bai[2] criticize the quality of our nuclear magnetic resonance (NMR) data and atom-by-atom analysis of global 'downhill' folding[3], also claiming that the data are compatible with two-state folding.

Rather than subjectively deciding which curves to use or reject, as favoured by Ferguson *et al.*[1], we selected chemical-shift data according to an objective criterion: namely, the approach to random-coil values at increasing temperature. We also used circular dichroism as an external reference. The melting temperatures were obtained consistently (and without normalization) by two independent methods: two- or three-state fits, with physically constrained baselines and derivative analysis[3]. Unconstrained two- or three-state fits produce essentially identical results (mean discrepancy with derivative method of ±3.7 K), indicating that baseline effects are minimal in our analysis. Crucially, the 54 curves with the lowest melting-temperature errors (mean standard error, ± 0.8 K) maintain the broad spread of melting temperatures (~50 K versus ~60 K in

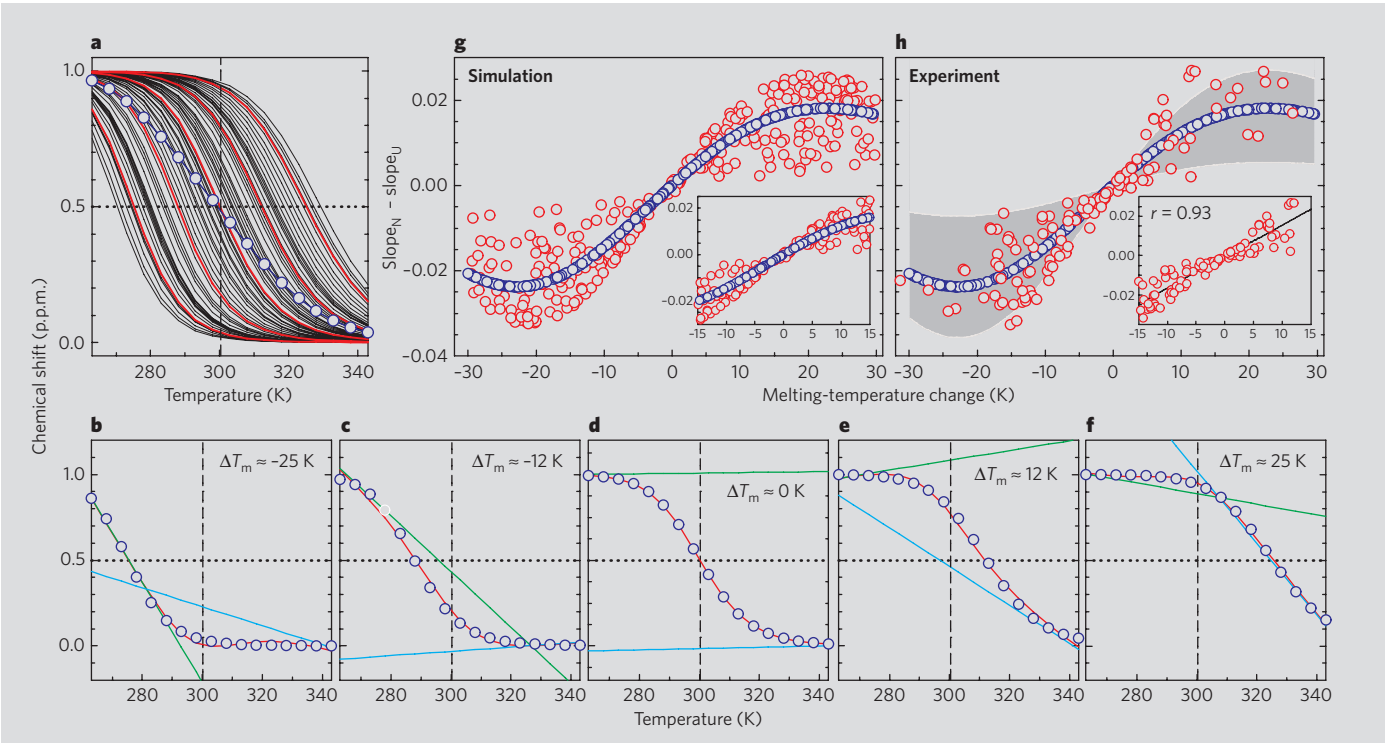


Figure 1 | Analysis of the physical relevance of global two-state fits. a, Fifty of the 300 synthetic curves generated by using a change in enthalpy of 90 kJ mol^{-1} and random melting temperatures (T_m) ranging from 270 to 330 K. The five curves also shown in **b–f** are in red; the average curve is in blue. **b–f,** Data (blue circles), fit (red line) and baselines for native (green) and unfolded (cyan) obtained in a global two-state fit. Vertical dashed lines indicate the fitted T_m; crossing of data with horizontal dotted lines shows the real T_m. **g,** Plot of the change in T_m versus the difference in native (N) and unfolded (U) baseline slopes. Blue circles, simulation with constant change in enthalpy of 90 kJ mol^{-1}; red circles, simulation with random values for both T_m and change in enthalpy (from 30 to 180 kJ mol^{-1}). Inset, quasi-linear region. **h,** Plot as in **g,** but with red circles showing the comparison between individual fits and global two-state fit for the 122 two-state-like proton unfolding curves of Naf-BBL. Shading indicates area occupied by red circles in **g**. Inset, quasi-linear region and linear regression analysis.

the whole data set). Furthermore, two rigorous statistical tests showed that our analysis reproduces exactly the first and second moments of the ensemble of NMR data[3]. The arguments by Ferguson *et al.*[1] and by Zhou and Bai[2] ignore or misinterpret this information.

In disagreement with their interpretations[1,2], global two-state fits demonstrate unambiguously that our NMR data are incompatible with two-state folding. A simple statistical *F*-test, which compares the ratio between fitting residuals and number of degrees of freedom of the statistically simpler (global two-state) model and the 158 individual fits, gives a probability of less than 0.0001 that the data originate from a two-state process. A global three-state model would perform even more poorly because it has nearly as many fitting parameters as the individual fits. Neither group[1,2] tests the statistical performance of their fits.

Furthermore, the global two-state fits to Naf-BBL are not physical, showing serious baseline manipulations. This is demonstrated in Fig. 1, which shows that global two-state models also fit 300 synthetic two-state curves with a melting enthalpy of 90 kJ mol^{-1} and randomly chosen melting temperatures spanning 60 K (Fig. 1a). The fitting is good, with residuals being less than 0.04% of the total amplitude, and fitted melting temperature of 300.3 ± 0.17 K. Obviously, this accurately determined global melting temperature is physically meaningless. The unique melting temperature is produced artificially by letting the baselines absorb all variability. Figure 1b–f illustrates this point. Baselines in Fig. 1d are almost horizontal because real and fitted melting temperatures are very close. Figure 1c, e shows baselines that compensate for lower (or higher) melting temperatures by crossing at temperatures above (or below) the globally fitted melting temperature. When the real and globally fitted melting temperatures are more than 20 K apart (Fig. 1b, f), the fits become physically absurd (although still quite good), with one real baseline becoming the apparent transition.

Correlations between baseline crossings and the discrepancies between individual and globally fitted melting temperatures are trademarks of unphysical baseline compensations. For data with constant melting enthalpy (for example, the data shown in Fig. 1a), such correlation produces a sinusoidal curve (Fig. 1g, blue circles). When the synthetic curves vary both in melting temperature and in melting enthalpy, as happens in Naf-BBL experiments[3], the curve becomes a bowtie-like distribution (Fig. 1g, red circles) because individual baselines must compensate for uncorrelated differences in melting temperatures and melting enthalpy.

Globally fitted baselines for the 122 two-state-like curves of Naf-BBL show exactly this behaviour (Fig. 1h), indicating that Naf-BBL is as two-state compliant as the set of synthetic curves shown in Fig. 1a. Baseline crossings (as in Fig. 1b–f) are evident in Fig. 1 of Zhou and Bai[2]. Ferguson *et al.*[1] do not show their baselines, although this is central to their argument. In fact, the fifteen ^{13}C curves claimed by these authors to demonstrate two-state folding in QNND-BBL[1,4] produce the same pattern as that shown in the inset of Fig. 1h, in their case with correlation coefficient $r = 0.97$ and a range of 18 K in melting-temperature differences. Therefore, the unfolding data from various BBL variants are consistent with one another and incompatible with two-state folding, as pointed out previously[3,5].

One-state (global downhill) folding offers a simple and straightforward quantitative explanation for the complex unfolding patterns discussed in refs 1–5. Conversely, Ferguson *et al.* and Zhou and Bai invoke a combination of fraying native state, intermediate and structured denatured state that merge with one another as judged by the crossing baselines, yet are separated by high-energy barriers[1,2]. Their interpretation is undermined according to Ockham's razor: it is more complex (several structurally varying states instead of one), only qualitative, physically implausible (requires infinitely narrow barriers), and provides no physical insight.

We have described the atom-by-atom analysis of downhill folding without mentioning its biological significance[3]. The concerns raised by Ferguson *et al.*[1] about this issue have been addressed elsewhere[5] and so are irrelevant here.

Mourad Sadqi, David Fushman, Victor Muñoz
Department of Chemistry and Biochemistry, and Center for Biomolecular Structure and Organization, University of Maryland, College Park, Maryland 20742, USA
e-mail: ymunoz@umd.edu

1. Ferguson, N., Sharpe, T. D., Johnson, C. M., Schartau, P. J. & Fersht, A. R. *Nature* **445,** doi: 10.1038/nature05643(2007).
2. Zhou, Z. & Bai, Y. *Nature* **445,** doi: 10.1038/nature05644 (2007).
3. Sadqi, M., Fushman, D. & Muñoz, V. *Nature* **442,** 317–321 (2006).
4. Ferguson, N. *et al. J. Mol. Biol.* **353,** 427–446 (2005).
5. Naganathan, A. N., Perez-Jimenez, R., Sanchez-Ruiz, J. M. & Muñoz, V. *Biochemistry* **44,** 7435–7449 (2005).

doi:10.1038/nature05645

ARTICLES

The architecture of human kin detection

Debra Lieberman[1,2], John Tooby[1] & Leda Cosmides[1]

Evolved mechanisms for assessing genetic relatedness have been found in many species, but their existence in humans has been a matter of controversy. Here we report three converging lines of evidence, drawn from siblings, that support the hypothesis that kin detection mechanisms exist in humans. These operate by computing, for each familiar individual, a unitary regulatory variable (the kinship index) that corresponds to a pairwise estimate of genetic relatedness between self and other. The cues that the system uses were identified by quantitatively matching individual exposure to potential cues of relatedness to variation in three outputs relevant to the system's evolved functions: sibling altruism, aversion to personally engaging in sibling incest, and moral opposition to third party sibling incest. As predicted, the kin detection system uses two distinct, ancestrally valid cues to compute relatedness: the familiar other's perinatal association with the individual's biological mother, and duration of sibling coresidence.

For the past 50 years, evolutionary biologists have argued that genetic relatedness should have played a role in the social evolution of species, such as humans, in which close genetic relatives frequently interact[1,2]. According to kin selection theory, computational variants that allocate altruistic effort effectively with respect to kinship outcompete variants that fail to regulate behaviour conditionally in response to relatedness. The effects of relatedness have been documented in a great diversity of taxa, ranging from social amoebas[3], social insects[4–6] and shrimp[7], to birds[8], aphids[9], plants[10,11], rodents[12] and primates[13–15]. To regulate behaviour conditionally in response to different degrees of kinship, organisms require mechanisms to discriminate genetic relatedness. Such mechanisms have been discovered in a variety of nonhuman species[16–18].

Equally, in long-lived, low-fecundity species with an open breeding structure (such as humans), the fitness of offspring is strongly affected by how closely parents are related. In such species, conceptive sexual behaviour between close genetic relatives produces offspring that suffer from inbreeding depression—a decline in fitness caused by rendering more deleterious recessives homozygous[19–21], and aggravated by parasites targeting more genetically homogeneous sets of hosts[22,23]. Consequently, heritable variants that cost-effectively reduce inbreeding depression by avoiding mating with close genetic relatives outcompete variants in which mating decisions are unaffected by relatedness.

The socioecology and population biology of human foragers[24–26] suggest that our ancestors would have been subject both to inbreeding depression and kin selection. This leads to the prediction that humans have an evolved system for detecting genetic relatedness, coupled to two output systems: one regulating altruism, the other regulating mate choice. Yet, there has been little research into the existence and design of human kin detection mechanisms[27–32].

The best-known exceptions are a handful of anthropological studies testing Westermarck's prescient 1891 hypothesis[33] that mutual exposure during childhood weakens sexual attraction among adults. These documented that non-relatives raised together in exceptional developmental circumstances (for example, crèche-mates or children cohabiting with future spouses) show lower rates of marriage or marital fertility, and higher rates of divorce and infidelity—archivally derived sociological measures used as proxies for the intensity of sexual desire[34,35]. But to map the information-processing architecture of a system predicted to detect genetic relatedness—and see whether it regulates altruistic as well as sexual motivation—it is necessary to measure the responses of living individuals drawn from a more species-characteristic range of family compositions, such as those that include actual genetic relatives.

Accordingly, the goal of the studies reported here was to test for the existence of a human kin detection system, and to test a series of basic predictions about its design features and architecture. It is ethically unacceptable to subject humans to the life-changing experimental manipulations used to discover kin detection systems in other species. So the architecture was mapped by quantitatively matching individual variation in the two predicted output systems—sibling altruism and opposition to incest—to naturally generated individual variation in developmental parameters that were predicted to serve as cues of relatedness.

Model of architecture and predictions

We propose that, for each familiar individual, *i*, the kin detection system computes and updates a continuous variable, the kinship index, KI_i, that corresponds to the system's pairwise estimate of genetic relatedness between self and *i*. These computational elements are regulatory variables that serve as input to neural programs regulating altruism towards *i* and, separately, to programs regulating sexual behaviour towards *i*.

Because relatedness cannot be directly observed, the system must be designed to register cues relevant to determining relatedness. To compute the kinship index, the system requires (1) monitoring circuitry designed to register cues to relatedness, and (2) a computational device, the kinship estimator, whose procedures have been tuned by a history of selection to take these registered inputs and transform them into a kinship index.

The cues the system uses cannot simply be derived ontogenetically ('learned') by identifying which arbitrary and transient cues happen to best predict relatedness in the local environment. To do this, the system would have to already know the relatedness of others—the very problem it needs to solve. Instead, the kin detection system must contain within its evolved design a specification of the core cues that it will use to determine relatedness—cues that reliably tracked genetic relatedness in the ancestral social environments that selected for the kin detection system.

[1]Center for Evolutionary Psychology, University of California Santa Barbara, Santa Barbara, California 93106, USA. [2]Department of Psychology, University of Hawaii, Honolulu, Hawaii 96822, USA.

For human foragers, a potentially informative cue to kinship is provided by the close perinatal association between mother and neonate that begins with birth and is enforced by the exigencies of early mammalian maternal care. Maternal perinatal association (MPA) provides a basis for the reliable mutual detection of mother and offspring and can, in turn, be used as an anchor point for sibling detection. Ancestrally, if an individual observed an infant in a durable, perinatal association with the individual's mother, then it was highly probable that that infant was the individual's sibling. We therefore proposed that sibling detection includes a monitoring subsystem specialized for registering MPA.

Although MPA is likely to be the single most informative cue, it cannot be used (for example) by younger siblings, because they are not alive at the time their older siblings are born and nursed. When MPA is unavailable, the kinship estimator should fall back on other cues that were highly predictive ancestrally. We predicted that the kin detection system would include a second subsystem specialized for registering the cumulative duration of coresidence summed over the full period they receive parental care. Ancestrally, parents (especially mothers) maintained close association with their children to care for them, and for this reason siblings co-associate statistically more than non-siblings. (Indeed, given the fusion–fission pattern of hunter–gatherer association, this same variable should—to some extent—link progressively more distant genetic relatives to increasingly diluted motivational residues.) Among human foragers, the maintenance of parental proximity for care delivery begins with birth and tapers off in late adolescence, a time when offspring become nearly independent adult foragers and when mating motivates new patterns of co-association[36,37]. Although this hypothesis differs from the ethological proposal of a period of early childhood imprinting[35], it is consistent with evidence that suggests that familiarity is a cue mediating kin detection in non-human primates[14,15,38,39].

The kinship estimator consists of algorithms for transforming the registered cues into the kinship index, a variable whose magnitude tracks relatedness between self and other. If the cues are integrated into a single index, then we should find that the same patterns of inputs are associated with the same patterns of outputs for both altruism and sexual aversion. This model (summarized in Fig. 1) leads to the following predictions.

(1) When MPA is absent, coresidence duration before adulthood with an individual should (a) upregulate altruism towards that individual, (b) upregulate sexual aversion towards that individual, and, as a by-product, (c) upregulate moral opposition[28,29] to third-party sibling incest.

(2) When MPA is present, it should produce the same three effects.

Selection should have tuned the procedures in the kinship estimator to use MPA and coresidence in a way that takes account of their relative informativeness and availability. Because MPA is the more robust, higher quality cue, we expect that when both are available, coresidence will be weighted by the kinship estimator far less than MPA, and perhaps not at all. Therefore, we propose a third prediction.

(3) When MPA is present, coresidence duration will not be as strong a predictor of altruistic motivations and sexual aversions. That is, the kinship estimator will use MPA in preference to coresidence duration in computing kinship.

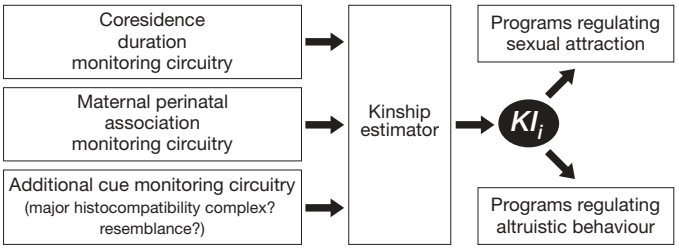


Figure 1 | Proposed model of the computational architecture of sibling detection. Cues to kinship are registered by cue monitoring circuits, which deliver their outputs to a kinship estimator. The kinship estimator uses these cues to compute the magnitude of a regulatory variable—a kinship index—for each individual, *i*, who is a potential sibling. The kinship index feeds into programs that regulate sibling altruism and sexual aversion.

Empirical investigation

Multiple, converging tests involving over 600 subjects were employed to assess whether particular developmental parameters (including MPA and coresidence duration) serve as cues to kinship and regulate both kin-directed altruism and sexual avoidance. Participants responded to questions regarding family composition and sibling interactions and were asked to complete instruments measuring: (1) frequency of altruistic behaviours towards a given sibling; (2) the intensity of altruistic motivation towards a given sibling; (3) the level of disgust evoked by the prospect of engaging in sexual acts with a given sibling, and (4) how morally wrong they perceive sibling incest among third parties to be (an unobtrusive measure of sexual aversion towards siblings[28,29]).

Results

The most important findings are displayed in Figs 2 and 3, which show that each of the two predicted cues of genetic relatedness for siblings—coresidence duration and maternal perinatal association—regulate outputs from the two functionally independent motivational systems (altruism and incest aversion) in the predicted way. (see Supplementary Information section 1).

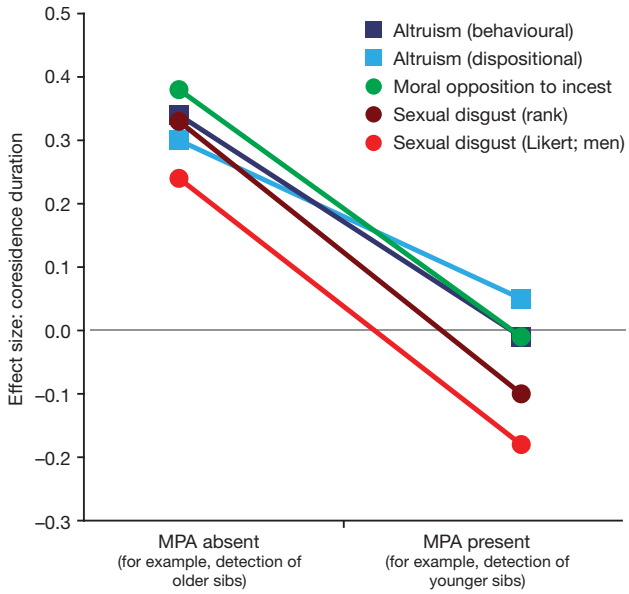


Figure 2 | Converging evidence indicates that the same computational variable, the kinship index, regulates disparate kin-relevant behaviours. The *x*-axis divides subjects into two groups—those who observed their mothers caring for their sibling as a neonate (MPA cue present) and those who did not (MPA cue absent). The *y*-axis shows the size of the correlation between coresidence duration and each dependent measure. Duration of coresidence predicts, with similar effect sizes, altruism and sexual aversions only when the cue of maternal perinatal association (MPA) is absent, as it is when younger siblings are detecting older ones. When the MPA cue is present, coresidence duration fails to predict sibling directed behaviours. This pattern appears for all measures: behavioural altruism, dispositional altruism, sexual disgust and moral judgments of sibling incest. Adaptive regulation of two distinct motivational output systems by the same pattern of inputs implicates a common underlying regulatory variable (see also Supplementary Information section 7).

The overall pattern of results was the same for men and women. For this reason, results are reported for both sexes combined, unless otherwise specified (see Methods).

When MPA is absent. When the MPA cue is absent—as is true whenever youngers are detecting older siblings—coresidence duration significantly predicts altruistic motivations and, separately, opposition to first and third person incest (Fig. 2). Subject's duration of coresidence with a particular sibling was positively correlated with all outcome measures: how much the subject helps that sibling (altruism: behavioural, $P = 6 \times 10^{-7}$ (or 8×10^{-7}, see Methods and Supplementary Information section 9), $N = 185$; dispositional, $P = 7 \times 10^{-6}$ (9×10^{-6}), $N = 185$); how disgusted the subject is at imagining sexual contact with that (opposite sex) sibling (sexual disgust (rank), $P = 0.0002$ (0.0003), $N = 114$; sexual disgust (Likert; men), $P = 0.0007$ (0.0009), $N = 156$, see Methods); and how morally wrong the subject judges third party sibling incest (moral opposition to incest, $P = 0.003$ (0.004), $N = 47$; see also refs 28, 29). Figure 2 shows that the effect sizes (r) for coresidence are very similar across widely divergent outcome variables, as would be expected if separate systems for altruism and sexual aversion were being regulated by the same internal variable, a kinship index.

When MPA is present. When the MPA cue is present—which can only be true for olders detecting younger siblings—levels of altruism and sexual aversion are high (Supplementary Information section 1). But in the presence of MPA, coresidence duration no longer predicts a single outcome measure (effect sizes ~0; Fig. 2; Supplementary Information section 2). Directed univariate analyses show that MPA and coresidence interact (sexual disgust (Likert; men), $P = 0.02$; sexual disgust (rank), $P = 0.003$; altruism (see Methods), $P = 0.03$; moral opposition, $P = 0.12$; see Supplementary Information section 1); the dramatic drop in effect sizes (all significant; Supplementary Information section 2) seen in Fig. 2 demonstrates that coresidence duration robustly affects altruism and sexual aversion in the absence, but not in the presence, of the MPA cue.

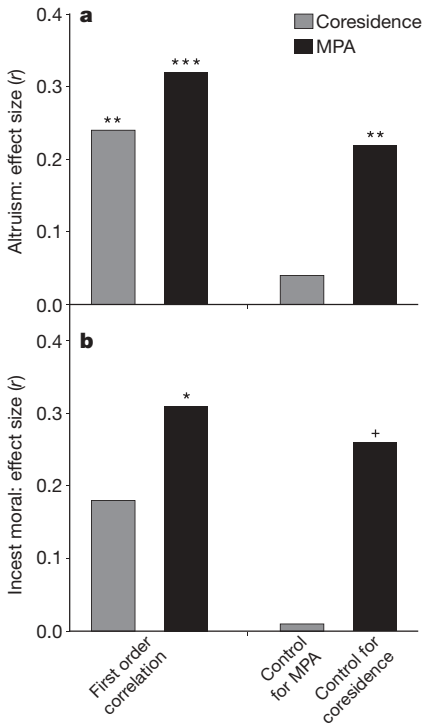


Figure 3 | When MPA and coresidence duration cues are both available, the kin detection system defaults to MPA, the more reliable cue. a, **b**, The only individuals for whom these cues could be jointly available are olders detecting younger siblings; each bar on the graph shows the size of the correlation between a cue and an outcome measure for this group. For olders responding to youngers, exposure to the MPA cue predicts both altruism (**a**) and moral opposition to sibling incest (**b**), and with the same effect size (black bar, first pair, each panel). The MPA cue continues to predict these disparate measures even after the effects of coresidence duration are statistically removed (black bar, second pair, each panel). In contrast, coresidence duration ceases to predict either altruism or moral opposition to sibling incest once the effects of MPA are removed (grey bar, second pair, each panel). ***$P < 0.001$, **$P < 0.01$, *$P = 0.05$, +$P < 0.10$.

MPA versus coresidence. MPA can only be observed by older siblings, and so they are the only individuals who can potentially be exposed to both MPA and coresidence duration cues. Thus analysis of olders allows one to see how the kinship estimator integrates these two cues to genetic relatedness.

Because MPA (as operationalized on these tests) is a dichotomous variable (1, 0) with 84% of older siblings scoring 1, its effects are most sensitively detected by using those outcome variables that are continuous and with high variance: altruism and moral opposition. The study assessing altruism yielded the most subject-and-younger sibling pairs ($N = 128$). As Fig. 3 shows, MPA significantly predicted altruism towards younger siblings ($r = 0.32$, $P = 0.0001$ (0.00013)), even when controlling for coresidence (partial $r = 0.22$, $P = 0.006$ (0.008), tolerance, 0.56, that is, much greater than the 0.10 collinearity threshold). This is important, because MPA and coresidence duration are themselves correlated ($r = 0.66$). In contrast, the relationship between coresidence duration and altruism towards younger siblings ($r = 0.24$) disappears when the effects of MPA are partialled out (partial $r = 0.04$, $P = 0.33$ (0.41)). When MPA, coresidence and beliefs about sibling kinship were all entered into a multiple regression, MPA was the only variable to independently predict variance in altruism towards younger siblings (partial $r = 0.27$, $P = 0.001$ (0.0013); tolerances, 0.42, 0.54 and 0.50, respectively). Moreover, MPA predicts altruism towards younger siblings better than either of its component parts (having the same mother + sibling coresidence beginning at the sibling's birth; see Supplementary Information section 3).

Although the sample size was much smaller ($N = 30$), the same MPA–coresidence pattern emerged for the moral wrongness judgments for incest (Fig. 3). For subjects with one opposite sex younger sibling, MPA predicted moral opposition at $r = 0.31$ ($P = 0.05$ (0.06)), about the same effect size as for altruism. When the effects of coresidence were statistically removed, the effect size for MPA remained virtually unchanged: $r = 0.26$. In contrast, the effect size for coresidence in predicting moral opposition was low ($r = 0.18$, $P = 0.17$ (0.21)), and when the effects of MPA were statistically removed, it disappeared entirely ($r = 0.01$, $P = 0.49$ (0.61); tolerance, 0.66).

Taken together, these analyses indicate that MPA is indeed a cue used by olders in detecting younger siblings; when MPA is present, coresidence duration is no longer used.

Alternative hypotheses

Is coresidence a kin cue or an artefact? When MPA is absent, coresidence duration correlates with altruism to the same degree regardless of the sibling's sex, as kin selection theory predicts that a cue to genetic relatedness should. But individuals are at risk for incest only from opposite sex siblings. Tellingly, moral opposition to third party sibling incest tracks duration of coresidence with an opposite sex, but not a same sex sibling ($r = -0.01$, $P = 0.47$ (0.59), $N = 30$). This pattern rules out any counter-hypothesis that coresidence duration is important not because it cues genetic relatedness, but because it is a spurious correlate of something else about the family (stability, traditional family structure, religion, and so on)[28].

The effects of coresidence when MPA is absent are also much targeted: duration of coresidence does not predict generosity outside of the sibling pair, and it is not positively correlated with moral judgments about any surveyed behaviours unrelated to incest (Supplementary Information section 4).

Early imprinting? Despite claims for an early imprinting period for sexual aversions[34,35], when MPA is absent, total duration of coresidence predicts altruism and sexual aversion better than age of sibling (or subject) when coresidence begins (start age; Supplementary Information section 5). The discovery that MPA is a potent cue for olders detecting younger sibs might explain past results suggesting an early imprinting period: start age at sibling's birth is not an independent predictor for olders detecting youngers (Supplementary Information section 3), but it is one component of the MPA cue.

Do beliefs matter? Coresidence duration predicted the outcome measures better than subjects' consciously held beliefs about siblings' genetic relatedness. Controlling for beliefs, coresidence continued to predict most outcome measures; in contrast, beliefs failed to predict most measures once the effects of coresidence were controlled for statistically (Supplementary Information section 6). Indeed, when subjects believe their sibling is step or adoptive, coresidence predicts altruism and sexual aversions, indicating that when beliefs conflict with the kin detection system, the criteria used by the kin detection system prevail (Supplementary Information section 6).

Other alternatives? Caution is always warranted in interpreting correlational findings, but it seems safe to say that altruism and sexual aversion are either regulated by the theoretically predicted cues, MPA and coresidence duration, or by unidentified cues very highly correlated with them. So far, we have been unable to find any cues that predict outcomes better than do MPA and coresidence duration.

Conclusions

The tight mesh between theoretical expectations and empirical tests provides strong support for the hypothesis that humans have a system designed by selection to detect genetic relatedness: specifically, one with (at a minimum) the computational elements outlined in Fig. 1. For example, the fact that different motivational systems are regulated in parallel by the same cues to relatedness implicates a single underlying neurocomputational variable—a kinship index—used by both. Moreover, if registered information about MPA and coresidence were fed directly into programs regulating altruism and sexual aversion, their effects would only be additive. They were not. Instead, the presence of MPA eliminated effects of coresidence. This is strong evidence for the existence of an intermediate computational device, the kinship estimator, equipped with procedures that combine these cues in a non-compensatory[40] way to compute the kinship index. These results contribute to a growing body of findings showing that humans are not immune to the evolutionary forces that have shaped other species, and that Darwinism has a central role in discovering the neural and psychological architecture of our species.

METHODS

All subjects completed a survey about family composition and attributes. For each sibling, subjects indicated that sibling's age, type of sibling (for example, biological, step), coresidence duration, age range of coresidence, and certainty of sharing the same biological mother and father[28]. From these, the following predictor variables were constructed: coresidence (duration of time a subject co-resided with his/her sibling between the subject's ages of 0 and 18); and Maternal Perinatal Association (MPA; where a score of 1 means the subject began coresidence with a sibling at the sibling's birth and is certain they share the same biological mother, and a score of 0 means any other scenario).

Instrument 1: sibling-directed altruism. Subjects ($N = 154$ (107 women); ages, 16–21, mean age $\pm$ s.d. of 18.44 ± 0.82; 287 sibling pairs) indicated the number of favours they performed for each sibling in the last month (behavioural measure), and, separately, how willing they would be to donate a kidney to their sibling (dispositional measure) on a 7-point Likert-like scale (0, not willing at all; 6, extremely willing). Responses from these measures produced the same pattern of results (Fig. 2) and were summed to produce a dependent variable, altruism (range, 0 to 16; mean $\pm$ s.d. of 7.57 ± 2.83).

Instrument 2: moral wrongness associated with third party sibling incest. Subjects ($N = 186$ (102 women); ages 18–47, mean $\pm$ s.d. of 21.54 ± 4.21) ranked 19 social transgressions on moral wrongness[28]. Two acts regarding third party sibling incest ('consensual sex between a brother and sister' and 'brother–sister marriage') were summed to produce a dependent variable, moral opposition (reverse-coded; range of 7 to 31 (mean $\pm$ s.d. of 22.43 ± 5.12)). This variable measures how morally wrong subjects view sibling incest among third parties (not incest with a particular sibling); therefore, to isolate effects to a particular sibling (in contrast to analyses in ref. 28), data analysis was restricted to individuals with only one opposite sex sibling ($N = 74$).

Instrument 3: disgust imagining sexual acts with a sibling (Likert). Subjects ($N = 455$ (264 women); ages 18–54, mean $\pm$ s.d. of 21.28 ± 3.91; a subset also completed Instruments 2 and 4) were asked how disgusting they would find engaging in various sexual and nonsexual behaviours on a 7-point Likert-like scale (0, not disgusting at all; 6, extremely disgusting). Among these were sexual acts with particular opposite sex siblings. For each opposite sex sibling, independent ratings for passionately kissing, and having sex with 'your sibling' were summed to produce a dependent variable, sexual disgust (Likert).

Initial analyses, for which non-independence was not a concern (see Supplementary Information section 8), indicated that women were at ceiling for this measure and showed significantly less variance than men in their responses (Levine's $F_{1,618} = 45.40$, $P = 4 \times 10^{-11}$). The multi response permutation procedure (MRPP)[41,42] indicated that, as predicted, women reported more disgust at sex with a sibling than did men (women (mean $\pm$ s.d.) 11.72 ± 0.98, $N = 264$; men 11.12 ± 1.96, $N = 191$; standardized test statistic of -12.72, $P = 5 \times 10^{-6}$). For this reason, this variable permitted the exploration of disgust responses in males, but not females ($N = 191$ males; ages 18–54, mean $\pm$ s.d. of 21.09 ± 3.30; 246 sibling pairs).

Sexual disgust (Likert) was transformed into a dichotomous variable: '1' was assigned if a male responded at ceiling for disgust associated with sex and kissing a sibling; '0' if otherwise (mean = 0.73, s.d. = 0.45). For the other three dependent measures, there were no sex differences in the relationships between predictor and outcome variables so results are reported for men and women together.

Instrument 4: disgust imagining sexual acts with a sibling (rank). A subset of participants who completed Instrument 3 also completed Instrument 4 ($N = 375$), which asked participants to assign a unique rank of disgust from 1 (not disgusting at all) to 50 (extremely disgusting) to eight acts, some of which involved sexual contact with a family member, short of intercourse. Using the rank of the sexual act involving a sibling, a variable, sexual disgust (rank), was constructed (women, mean = 47.36, s.d. = 3.99; men, mean = 45.51, s.d. = 9.91). To assess the effects of coresidence on sexual disgust in a way that reflects coresidence with a particular sibling, data analyses are limited to subjects with only one opposite sex sibling ($N = 243$ (144 women); ages 18–50, mean $\pm$ s.d. of 21.02 ± 2.95).

Data analyses. Correlations involving dependent measures 'moral opposition' and 'sexual disgust' controlled for the subject's sexual orientation. Controlling for social desirability yielded similar effect sizes. For univariate analyses, we used directed tests to assess predicted effects[43]. Pearson correlations for which we had prior predictions report one-tailed *P*-values, followed by directed *P*-values in parentheses (see Supplementary Information section 9). Non-independence occurs in Instruments 1 and 3 because some subjects have multiple siblings thus contributing multiple data-points. For these two studies, separate analyses using only one sibling pair per subject were carried out and yielded the same effect sizes (see Supplementary Information section 8).

Received 22 July; accepted 5 December 2006.

1. Hamilton, W. D. The genetical evolution of social behaviour. I, II. *J. Theor. Biol.* **7**, 1–52 (1964).
2. Williams, G. C. & Williams, D. C. Natural selection of individually harmful social adaptations among sibs with special reference to social insects. *Evolution* **11**, 32–39 (1957).
3. Strassmann, J. E., Zhu, Y. & Queller, D. C. Altruism and social cheating in the social amoeba *Dictyostelium discoideum*. *Nature* **408**, 965–967 (2000).
4. Crozier, R. H. & Pamilo, P. *Evolution of social insect colonies: Sex allocation and kin-selection* (Oxford Univ. Press, Oxford, 1996).
5. Chapuisat, M. & Keller, L. Testing kin selection with sex allocation data in eusocial hymenoptera. *Heredity* **82**, 473–478 (1999).
6. Passera, L., Aron, S., Vargo, E. L. & Keller, L. Queen control of sex ratio in fire ants. *Science* **293**, 1308–1310 (2001).
7. Duffy, J. E. Eusociality in a coral-reef shrimp. *Nature* **381**, 512–514 (1996).
8. Baglione, V., Canestrari, D., Marcos, J. & Ekman, J. Kin selection in cooperative alliances of carrion crows. *Science* **300**, 1947–1949 (2003).
9. Ito, Y. The evolutionary biology of sterile soldiers in aphids. *Trends Ecol. Evol.* **4**, 69–73 (1989).
10. Queller, D. C. Inclusive fitness in a nutshell. *Oxford Surveys Evol. Biol.* **6**, 73–109 (1989).
11. Cosmides, L. & Tooby, J. Cytoplasmic inheritance and intragenomic conflict. *J. Theor. Biol.* **89**, 83–129 (1981).
12. Sherman, P. W. Nepotism and the evolution of alarm calls. *Science* **197**, 1246–1253 (1977).

13. Buchan, J. C., Alberts, S. C., Silk, J. B. & Altmann, J. True paternal care in a multi-male primate society. *Nature* **425,** 179–181 (2003).
14. Chapais, B. & Berman, C. M. (eds) *Kinship and Behavior in Primates* (Oxford Univ. Press, New York, 2004).
15. Silk, J. B. Kin selection in primate groups. *Int. J. Primatol.* **23,** 849–875 (2002).
16. Fletcher, D. & Michener, C. (eds) *Kin Recognition in Animals* (Wiley, New York, 1987).
17. Hepper, P. G. *Kin Recognition* (Cambridge Univ. Press, New York, 1991).
18. Holmes, W. The early history of Hamiltonian-based kin recognition research theory: past and future. *Ann. Zool. Fennici* **41,** 691–711 (2004).
19. Charlesworth, B. & Charlesworth, D. The genetic basis of inbreeding depression. *Genet. Res.* **74,** 329–340 (1999).
20. Crnokrak, P. & Roff, D. A. Inbreeding depression in the wild. *Heredity* **83,** 260–270 (1999).
21. Bittles, A. H. & Neel, J. V. The costs of human inbreeding and their implications for variation at the DNA level. *Nature Genet.* **8,** 117–121 (1994).
22. Tooby, J. Pathogens, polymorphism, and the evolution of sex. *J. Theor. Biol.* **97,** 557–576 (1982).
23. Penn, D. J. & Potts, W. K. The evolution of mating preferences and major histocompatibility coupled genes. *Am. Nat.* **153,** 145–164 (1999).
24. Lee, R. B. & Devore, I. *Man the Hunter* (Aldine, Chicago, 1968).
25. Howell, N. *Demography of the Dobe! Kung* 2nd edn (Aldine Transaction, New York, 2000).
26. Hill, K. & Hurtado, A. *Ache Life History: The Ecology and Demography of a Foraging People* (Aldine Transaction, New York, 1996).
27. Bevc, I. & Silverman, I. Early separation and sibling incest: A test of the revised Westermarck theory. *Evol. Hum. Behav.* **21,** 151–161 (2000).
28. Lieberman, D., Tooby, J. & Cosmides, L. Does morality have a biological basis? An empirical test of the factors governing moral sentiments regarding incest. *Proc. R. Soc. Lond. B* **270,** 819–826 (2003).
29. Fessler, D. M. T. & Navarrete, C. D. Third-party attitudes toward sibling incest: Evidence for Westermarck's hypotheses. *Evol. Hum. Behav.* **25,** 277–294 (2004).
30. Wedekind, C. & Furi, S. Body odour preferences in men and women: do they aim for specific MHC combinations or simply heterozygosity? *Proc. R. Soc. Lond. B* **264,** 1471–1479 (1997).
31. Ober, C. *et al.* HLA and mate choice in humans. *Am. J. Hum. Genet.* **61,** 497–504 (1997).
32. DeBruine, L. M. Trustworthy but not lust-worthy: Context-specific effects of facial resemblance. *Proc. R. Soc. Lond. B* **272,** 919–922 (2005).
33. Westermarck, E. A. *The History of Human Marriage* 5th edn (Macmillan, London, 1891/, 1921).
34. Wolf, A. P. *Sexual Attraction and Childhood Association: A Chinese Brief for Edward Westermarck* (Stanford Univ. Press, Stanford, California, 1995).
35. Shepher, J. Mate selection among second generation kibbutz adolescents and adults: incest avoidance and negative imprinting. *Arch. Sex. Behav.* **1,** 293–307 (1971).
36. Kaplan, H. *et al.* A theory of human life history evolution: Diet, intelligence, and longevity. *Evol. Anthropol.* **9,** 156–185 (2000).
37. Hewlett, B. & Lamb, M. *Hunter–Gatherer Childhoods* (Aldine Transaction, Somerset, New Jersey, 2005).
38. Walters, J. R. in *Kin Recognition in Animals* (eds Fletcher, D. J. C. & Michener, C. D.) 359–393 (Wiley & Sons, New York, 1987).
39. Bernstein, I. in *Kin Recognition* (ed. Hepper, P. G.) 6–29 (Cambridge Univ. Press, Cambridge, 1991).
40. Gigerenzer, G., Todd, P., ABC Research Group. *Simple Heuristics That Make Us Smart* (Oxford Univ. Press, New York, 1999).
41. Mielke, P. W. & Berry, K. J. *Permutation Methods: A Distance Function Approach* (Springer, New York, 2001).
42. Cade, B. S. & Richards, J. D. *User Manual for BLOSSOM Statistical Software* (Midcontinent Ecological Science Center, US Geological Survey, Fort Collins, Colorado, 2005).
43. Rice, W. R. & Gaines, S. D. Heads I win, tails you lose: Testing directional alternative hypotheses in ecological and evolutionary research. *Trends Ecol. Evol.* **9,** 235–237 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors thank P. Boyer, D. Fessler, S. Gangestad, P. Pocker, H. Waldow, G. Williams, D. Williams, UCSB Academic Senate and the providers of the NSF Presidential Young Investigator Award (J.T.), and NIH Director's Pioneer Award (L.C.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.L. (debra@debralieberman.com).

ARTICLES

Structural definition of a conserved neutralization epitope on HIV-1 gp120

Tongqing Zhou[1], Ling Xu[1], Barna Dey[1], Ann J. Hessell[3], Donald Van Ryk[2], Shi-Hua Xiang[4], Xinzhen Yang[4], Mei-Yun Zhang[5], Michael B. Zwick[3], James Arthos[2], Dennis R. Burton[3], Dimiter S. Dimitrov[5], Joseph Sodroski[4], Richard Wyatt[1], Gary J. Nabel[1] & Peter D. Kwong[1]

The remarkable diversity, glycosylation and conformational flexibility of the human immunodeficiency virus type 1 (HIV-1) envelope (Env), including substantial rearrangement of the gp120 glycoprotein upon binding the CD4 receptor, allow it to evade antibody-mediated neutralization. Despite this complexity, the HIV-1 Env must retain conserved determinants that mediate CD4 binding. To evaluate how these determinants might provide opportunities for antibody recognition, we created variants of gp120 stabilized in the CD4-bound state, assessed binding of CD4 and of receptor-binding-site antibodies, and determined the structure at 2.3 Å resolution of the broadly neutralizing antibody b12 in complex with gp120. b12 binds to a conformationally invariant surface that overlaps a distinct subset of the CD4-binding site. This surface is involved in the metastable attachment of CD4, before the gp120 rearrangement required for stable engagement. A site of vulnerability, related to a functional requirement for efficient association with CD4, can therefore be targeted by antibody to neutralize HIV-1.

The human immunodeficiency virus type 1 (HIV-1) crossed from chimpanzees to humans early in the twentieth century and has since infected ~1% of the world's adult population[1,2]. This spread and the absence of an effective vaccine are to a large degree a consequence of the ability of HIV-1 to evade antibody-mediated neutralization[3–5]. On HIV-1, the only viral target available for neutralizing antibodies is the envelope spike, which is composed of three copies of the gp120 exterior envelope glycoprotein and three gp41 transmembrane glycoprotein molecules[6,7]. Genetic, immunological and structural studies of the HIV-1 envelope glycoproteins have revealed extraordinary diversity, manifest in a variety of immunodominant loops, as well as multiple overlapping mechanisms of humoral evasion, including self-masquerading glycan and conformational masking[8–11]. These evolutionarily honed barriers of diversity and evasion have confounded traditional vaccine development.

Two strategies have been proposed to surmount these barriers: examination of known broadly neutralizing antibodies (2F5, 2G12, 4E10 and b12) to identify susceptible targets of neutralization[12], and analysis of functional constraints to identify potential sites of vulnerability[13]. To facilitate viral entry, the gp120 glycoprotein must bind to cell-surface CD4 (ref. 14), alter its conformation to reveal a site for co-receptor attachment[15], and trigger conformational rearrangements in the gp41 glycoprotein to mediate fusion of viral and host cell membranes[16,17]. Constraints on envelope (Env) variation and exposure, associated with required functions of viral entry, provide potential footholds for broad, antibody-mediated neutralization.

Here we combine analysis of function with clues from antibody. We constructed stabilized gp120 molecules, constrained to stay in the CD4-bound conformation even in the absence of CD4, and tested the effect of this stabilization on the binding kinetics of CD4 and on antibodies reactive with sites of receptor binding. We then determined the crystal structure of the broadly neutralizing antibody b12 in complex with one of the stabilized gp120 molecules. Analysis of this structure, combined with detailed antigenic analyses of gp120 molecules stabilized to various extents in the CD4-bound conformation, not only reveals the functionally conserved surface that allows for initial CD4 attachment, but also provides an atomic-level description of the b12 epitope, which serves as a key target for humoral neutralization of HIV-1.

Recognition of the CD4-binding site

Conformational flexibility of HIV-1 gp120 complicated analysis of antibody recognition. To circumvent this complication, we used an iterative structure-based scheme to stabilize gp120 in its CD4-bound state (Table 1, Supplementary Fig. 1 and Supplementary Tables 1 and 2). The CD4-bound state of gp120 comprises an inner domain, outer domain and four-stranded bridging sheet mini-domain[18]. Five disulphides and four cavity-altering substitutions were created to restrict interdomain movements and to stabilize bridging sheet formation. Crystallographic analysis of gp120 variants with these substitutions in complex with CD4 and antibody 17b at 2.0–2.2 Å resolution showed that four of five disulphides formed and that disulphide and cavity-altering substitutions induced minimal structural perturbation. We also analysed a two-disulphide variant at 2.5 Å resolution as well as a three-disulphide variant at 2.8 Å resolution (Table 1). In both structures, all potential disulphides formed.

To measure the degree of conformational fixation, we used isothermal titration calorimetry to assess the entropy of interaction between the different stabilized gp120 cores and CD4 (Table 1 and Supplementary Table 3). Disulphides had a more substantial effect than cavity-altering substitutions. Tethering the centre two strands of the bridging sheet (by linking residues 123–431) or the termini-proximal ends of the domains (96–275 or 231–267) reduced the entropy of interaction by 15–30%. A more substantial effect (a 60% reduction) was gained by tethering the bridging sheet to the inner domain (109–428). Thermodynamic analysis thus quantified

[1]Vaccine Research Center, and [2]Laboratory of Immunoregulation, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. [3]Departments of Immunology and Molecular Biology, Scripps Research Institute, La Jolla, California 92037, USA. [4]Department of Cancer Immunology and AIDS, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. [5]Center for Cancer Research, National Cancer Institute, Frederick, Maryland 21702, USA.

Table 1 | Characterization of stabilized gp120 cores.

gp120*	Resolution (Å)	Entropy of CD4 interaction ($-T\Delta S$) (kcal mol^{-1})	Ligands															
			CD4		CD4-induced antibodies				CD4-binding-site antibodies									
					17b		m6		b12	b3	b6	b11	b13	m14	m18	F91	F105	15e
			On† ($\times10^4$)	K_d (nM)	On† ($\times10^4$)	K_d (nM)	On† ($\times10^4$)	K_d (nM)	K_d(WT)/K_d(mutant)									
Wild type (WT)	2.00	39.7	2.5	52	0.69	560	2.2	2,400	1	1	1	1	1	1	1	1	1	1
F2	2.05	39.1	4.7	3.5	2.9	530	14	680	0.28	1.0	4.6×10^{-5}	1.6×10^{-3}	0.73	5.2×10^{-3}	0.29	9.0×10^{-3}	$<10^{-5}$	$<10^{-5}$
Ds1 F123	2.20	28.3	3.5	14	3.4	230	12	490	0.43	0.29	1.9×10^{-5}	9.0×10^{-4}	0.95	0.059	0.21	4.2×10^{-3}	$<10^{-5}$	$<10^{-5}$
Ds2 F2	2.00	17.2	5.3	4.8	10	200	32	400	0.085	$<10^{-5}$	$<10^{-5}$	$<10^{-5}$	0.029	$<10^{-5}$	1.5×10^{-4}	$<10^{-5}$	$<10^{-5}$	$<10^{-5}$
Ds3 F2	2.00	26.9	3.8	1.9	9.5	120	25	220	0.42	0.72	3.3×10^{-5}	2.5×10^{-3}	0.67	3.3×10^{-3}	0.28	2.7×10^{-3}	$<10^{-5}$	$<10^{-5}$
Ds4 F2	2.00	34.2	4.5	2.8	2.3	820	8.4	2,200	0.30	0.97	0.49	1.6×10^{-3}	0.56	2.4×10^{-3}	0.34	6.2×10^{-3}	$<10^{-5}$	$<10^{-5}$
Ds12 F123	2.50	18.5	4.9	10	3.7	320	19	710	0.60	$<10^{-5}$	$<10^{-5}$	$<10^{-5}$	7.6×10^{-3}	$<10^{-5}$	3.2×10^{-3}	$<10^{-5}$	$<10^{-5}$	$<10^{-5}$
Ds123 F12	2.80	18.7	3.0	1.9	66	19	140	32	1.1	$<10^{-5}$	$<10^{-5}$	$<10^{-5}$	1.6×10^{-3}	$<10^{-5}$	2.6×10^{-3}	$<10^{-5}$	$<10^{-5}$	$<10^{-5}$

Stabilized gp120 cores were characterized by X-ray crystallography, isothermal titration calorimetry and surface-plasmon resonance. K_d, dissociation constant.
* gp120 core variants as defined by cavity-altering substitutions: F1, M95W; F2, T257S/S375W; F3, A433M; and also by disulphide (Ds)-bond-forming cysteine substitutions: Ds1, W96C/V275C; Ds2, I109C/Q428C; Ds3, T123C/G431C; Ds4, K231C/E267C.
† On-rate of ligand binding to gp120 (units in $M^{-1}s^{-1}$).

the degree of variant gp120 stabilization and demonstrated substantial remaining flexibility, despite the presence of up to three stabilizing disulphides.

Because the epitope for CD4-induced antibodies is only formed in the CD4-bound state[10,19,20], we could use the increase in antibody on-rate to assess the degree to which stabilization 'preformed' the variant gp120 molecules in the CD4-bound state. Reasonable correlations were observed between the on-rate for CD4-induced antibodies and the entropy of CD4 binding ($r = 0.74$, $P < 0.036$ for antibody 17b (ref. 19); $r = 0.76$, $P < 0.029$ for antibody m6 (ref. 21)), with the most extreme change for the three-disulphide variant (a 95-fold increase in on-rate for antibody 17b and a 65-fold increase for antibody m6) (Table 1 and Supplementary Fig. 1c).

Notably, almost no change in CD4 on-rate was observed with the stabilized gp120 molecules (Table 1, Supplementary Table 4 and Supplementary Fig. 1c), suggesting that the initial contact surface recognized by CD4 is present in unliganded gp120 before the CD4-induced conformational change. Similarly invariant on-rates have recently been reported for CD4 binding to gp120 molecules restrained from achieving a CD4-bound conformation by removing a flexible loop[22]. Thus, in contrast to the CD4-induced antibodies, a conformational change is not required to expose an initial site of contact for CD4. Rather, it serves to lock CD4 into place once contact has been made.

To determine the degree to which fixation of gp120 in its CD4-bound conformation restricts antigenic recognition, we tested binding of a panel of CD4-binding-site antibodies to the stabilized gp120 variants (Table 1). The average overall difference in free energy of binding for CD4-binding-site antibodies to wild type and stabilized variants correlated well with the change in entropy of variant gp120 binding to CD4 ($r = 0.89$, $P < 0.0035$) (Supplementary Fig. 1c). Conformational constraint thus effectively hides the CD4-bound conformation of gp120 from recognition by antibodies that target the site of CD4 binding. Indeed, of all the CD4-binding-site antibodies tested, only the unique, broadly neutralizing antibody b12 was able to bind well to stabilized gp120.

Structure of a b12–gp120 complex

Although gp120 is the primary target of neutralizing antibodies elicited during natural infection[23], most gp120-reactive antibodies are ineffective at neutralizing primary HIV-1 isolates (reviewed in refs 4, 24). Only two gp120-reactive antibodies (b12, 2G12) with effective neutralization activity against diverse primary HIV-1 isolates have thus far been identified[25–27]. *In vivo*, b12 can protect macaques against vaginal challenge from pathogenic simian–human immunodeficiency virus (SHIV)[28]. Although the structure of the entire, unbound b12 immunoglobulin (IgG1) has been determined[29], a molecular description of its interaction with gp120 remained elusive.

To facilitate a structural understanding of b12 neutralization of HIV-1, we screened crystallizations of b12 in complex with various forms of gp120. The lack of b12 conformational fixation of gp120 (ref. 10), coupled to inherent gp120 flexibility, complicated crystallization of a b12–gp120 complex; to overcome this complication, we

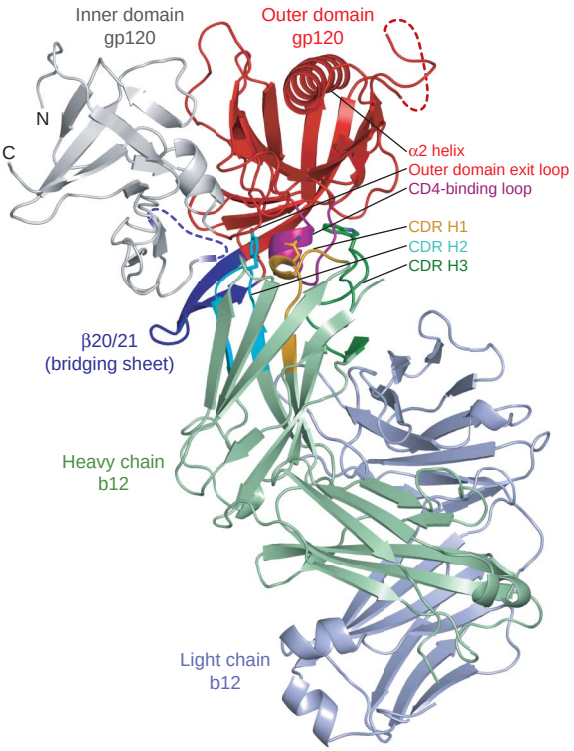


Figure 1 | Structure of b12 in complex with an HIV-1 gp120 core. Polypeptide chains are depicted in ribbon representation, with disordered regions as dashed lines. The gp120 inner domain is grey, and the outer domain is red, except for the CD4-binding loop, which is purple. The strands and associated loops, which in the CD4-bound conformation correspond to the bridging sheet, are blue. The b12 light chain is blue–grey and the b12 heavy chain is green, with associated CDRs highlighted in orange (H1), cyan (H2) and dark green (H3). The heavy-chain dominance of the binding interaction is apparent, with the nearest light-chain approach separated by ~10 Å from gp120. Heavy-chain-only interactions are rare, although heavy-chain interactions predominate in a number of viral Env–antibody complexes, including those from SARS coronavirus[49] and influenza virus haemagglutinin[50]. Three b12 residues (Asn 31, Tyr 53 and Trp 100) from each of the heavy-chain CDRs are depicted in stick representation. Together, these three residues combine to form ~40% of the b12 contact surface. They can be seen gripping the CD4-binding loop, the central focus of the b12 interaction with gp120.

used the mutationally stabilized gp120 molecules. Complexes with wild-type core or with single-disulphide variants (96–275 or 123–431) failed to produce crystals suitable for structural analysis. However, diffraction to 2.3 Å was obtained from hexagonal crystals of the antigen-binding fragment (Fab) of b12 in complex with a two-disulphide-stabilized gp120 core. We solved the complex structure by molecular replacement and refined it to an *R*-value of 19.3% (R_{free} = 25.5%) (Supplementary Table 5).

Notably, only the heavy chain of b12 interacted with gp120, with each of the three heavy-chain complementarity-determining regions (CDRs) making extensive contact (Fig. 1, Supplementary Table 6 and Supplementary Figs 2 and 3). Despite this unusual heavy-chain-only usage, the surface areas of interaction were in the range typical for antibody–protein interfaces (reviewed in ref. 30), with a total of 1,480 Å^2 buried[31] in the interaction (737 Å^2 on gp120 and 743 Å^2 on b12).

The gp120 surface bound by b12 was confined largely to the gp120 outer domain, consistent with predictions from alanine substitutions (Supplementary Table 7 and ref. 32). We examined this surface for features that might make it amenable to antibody recognition. The outer domain is composed of two barrels, stacked end to end (Fig. 2). The termini-distal barrel comprises seven anti-parallel β-strands; the other contains six β-strands embracing the α2 helix. b12 binds at the barrel–barrel juncture to the gp120 face opposite the α2 helix. Although most of the gp120 backbone is involved in secondary structure interactions, the barrel juncture contains a number of structural elements with unpaired backbone, often used by antibodies in recognition. Principal among these are two parallel loops—the CD4-binding loop and the outer domain exit loop—that extend from the only two parallel β-strands in the entire gp120 core. b12 takes advantage of this backbone reactivity, forming six direct and four water-mediated hydrogen bonds with the main-chain atoms of these two loops (Supplementary Fig. 2). Overall, the outer domain comprises 82% of the gp120 contact surface with b12, with the CD4-binding loop forming a little over one-half of this surface.

Comparison of b12- and CD4-binding

Both CD4 and b12 bind primarily to the outer domain of gp120, which is remarkably well preserved between unliganded, b12- and CD4-bound states (Figs 2 and 3). Their entropies of interaction, however, are very different, with CD4 inducing a 40–50 kcal mol^{-1} change, and b12 inducing only a 6 kcal mol^{-1} change[10]. This difference in conformational fixation of gp120 is seen in the divergent atomic-mobility values of the domains (Fig. 3b).

The angles of approach of CD4 and b12 to gp120 are similar, although not precisely the same (Fig. 3c). If the gp120 outer domains of b12- and CD4-bound structures were superimposed to orient equivalently b12 and CD4, about one-half of CD4 domain 1, which makes all of the contacts with gp120, is not encompassed by the b12 Fab (despite the volume of the Fab variable domains being twice as large as that of domain 1 of CD4). The projection of CD4 outside of the angle of b12 approach suggests that the parameters that sterically restrict b12 binding to the functional spike are not overly stringent.

The contact surfaces of CD4 and b12 on gp120 have considerable overlap (Fig. 3a and Supplementary Fig. 4). Most of this overlap is on the outer domain, where the CD4-binding loop is a central focus of binding for both CD4 and b12. CD4 and b12, however, interact with the CD4-binding loop quite differently (Fig. 3d). b12 uses all three of its CDR heavy-chain loops to grasp virtually all surface-exposed portions of the loop. In contrast, CD4 only binds to one side of the loop, making anti-parallel hydrogen bonds between CD4- and gp120-main-chain atoms.

The primary difference between b12 and CD4 interactions with gp120 involves the conformationally mobile β20/21. For b12, these interactions are peripheral to the binding surface, with alanine substitution of the primary b12 contact with β20/21 (at b12 residue Asn 56) having little impact on overall b12 binding[33]. In contrast, CD4 interactions with β20/21 form an integral part of the binding surface, burying 160 Å^2 of surface area and forming a topologically contiguous contact surface with the CD4-binding loop.

To delineate further the differences in binding between b12 and CD4, we characterized binding interactions with an HIV-1 gp120 fragment, termed OD1 (ref. 34). This fragment comprises residues 252–482 and encompasses the entire outer domain including V3 as well as the β20/21 excursion. Binding of b12 to OD1 showed nearly identical rates of association compared to binding of b12 to HXBc2 core gp120, although the dissociation rate was about 15-fold more rapid (Supplementary Fig. 5). We were unable to detect binding of CD4 to OD1 (Supplementary Fig. 5). Because CD4 demonstrated virtually no change in on-rate when tested on conformationally stabilized gp120 molecules, we turned to a dodecameric variant of CD4 (D1D2-Igαtp (ref. 35)), as avidity from multivalent binding provides an effective means by which to reduce off-rate. We observed that D1D2-Igαtp binds with virtually identical rates of association to both OD1 and core gp120 (Supplementary Fig. 5).

The results suggest the following series of molecular interactions for b12 and CD4 binding to gp120 (Supplementary Fig. 6). Initial contact by CD4 occurs with the structurally invariant outer domain, to a surface constitutively exposed on the envelope spike. In primary isolates, which are generally resistant to neutralization by soluble CD4, this interaction is not stable and CD4 readily 'falls off'. However, at the cell surface (or with dodecameric CD4), multiple CD4 molecules can bind simultaneously to the viral spike and use avidity to enhance stability. The avidity-enhanced outer domain–CD4 complex provides a receptive contact surface for the bridging sheet. A highly coordinated rearrangement of the inner domain allows for formation of the bridging sheet, which welds CD4 into place.

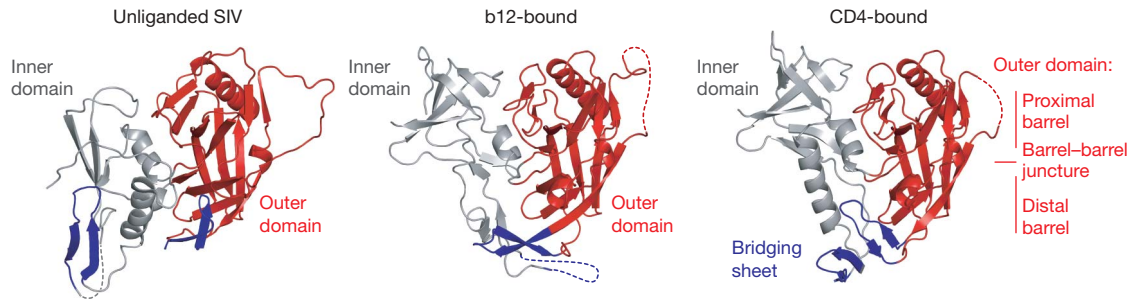


Figure 2 | Conformational states of gp120. The unliganded, b12- and CD4-bound conformations of gp120 are depicted, with polypeptide in ribbon representation and disordered regions as dashed lines. Inner domains are grey, outer domains are red and regions that in the CD4-bound state correspond to the bridging sheet are blue. Both b12- and CD4-bound conformations are of the Ds12 F123 variant of HIV-1, whereas the unliganded structure is of simian immunodeficiency virus (SIV). Comparison of these three gp120 conformations highlights not only the structural plasticity of the inner domain and bridging sheet, but also the conformational stability of the outer domain.

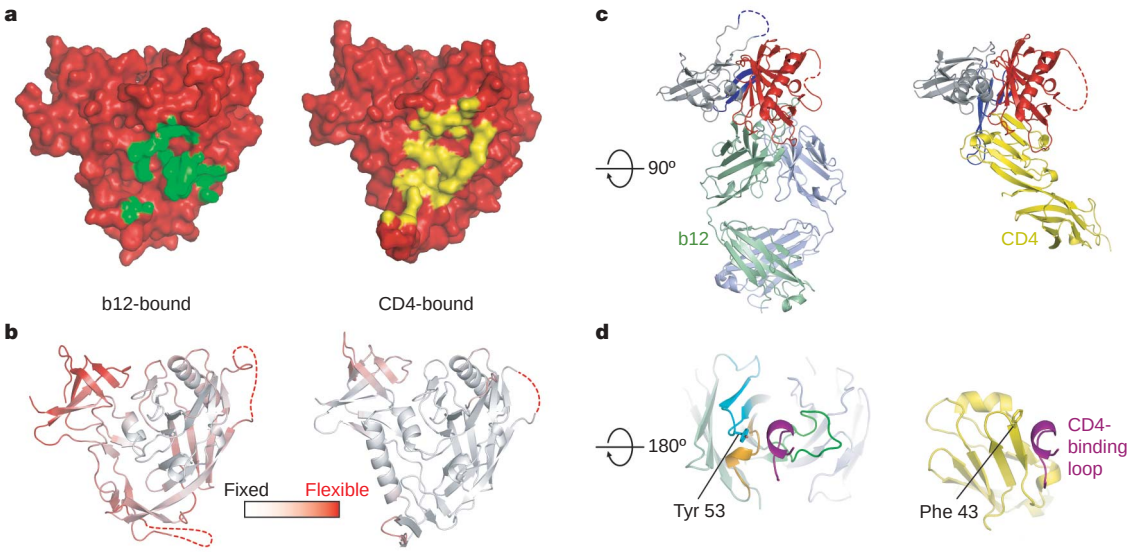


Figure 3 | b12 and CD4 recognition of gp120. The orientation of gp120 in **a** and **b** is the same as in Fig. 2, with **c** and **d** rotated about a horizontal axis by 90° and 180°, respectively. **a**, Molecular surface of gp120 in red, with the b12-contact surface in green (left) and the CD4-contact surface in yellow (right). **b**, Ribbon diagram of the b12- and CD4-bound gp120 coloured according to the atomic mobility of the polypeptide, with white for fixed and red for flexible. In the b12-induced conformation, only the outer domain is fixed by interaction with antibody, with the average atomic mobility of the outer domain about one-half that of the inner domain. By contrast, CD4 fixes the entire core, resulting in outer and inner domains of similar overall atomic mobility. **c**, Comparison of b12 and CD4 angles of attachment. The polypeptide chains are depicted in ribbon representation, and coloured according to Fig. 1, with CD4 in yellow. The similarity in angle of attachment as well as the binding focus of b12 on the outer domain (red) are evident in this orientation. **d**, The CD4-binding loop of gp120 (purple) is shown with b12 (left) and CD4 (right). There are parallels between several key CD4 contact residues and those of b12. For example, Phe 43 of CD4 (shown in stick model) inserts into a critical juncture at the nexus of the inner, outer and bridging sheet regions of gp120, whereas Tyr 53 of b12 (also shown in stick model) inserts at a similar position, although displaced by ~3 Å.

Contact by b12 occurs at the same constitutively exposed surface initially recognized by CD4. However, b12 is able to latch onto this outer domain surface with high affinity, without additional gp120 conformational change. This absence of conformational constraint allows b12 to bind and neutralize primary isolates that otherwise would be protected by conformational masking. In this manner, b12 uses the functionally conserved initial contact site for CD4 on gp120 to neutralize effectively HIV-1 (Fig. 4).

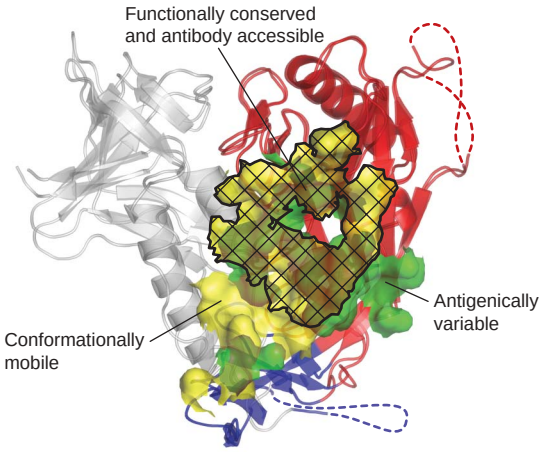


Figure 4 | Structural definition of a conformationally invariant, antibody-accessible portion of the CD4-binding site. The b12- and CD4-bound conformations of gp120 are shown in ribbon representation, after superposition of outer domains (red). A semitransparent molecular surface shows the contact surfaces of b12 (green) and CD4 (yellow). Subsets of these surfaces, corresponding to regions of conformational flexibility (for example, of the inner domain (grey) or bridging sheet (blue)), are delineated, as are regions of b12 contact outside of the conserved CD4-binding site. As can be seen, functional analysis serves to transcend the particulars of b12 binding, whereas antibody defines accessibility. Although we have formally shown only the b12 contact surface to be accessible in the context of a functional viral spike, the highly effective neutralization of D1D2-Igαtp and the kinetics of its association with both core and OD1 variants of gp120 suggest that the CD4-binding surface on the outer domain is accessible.

A site of HIV-1 vulnerability

Wedged between the glycan-shielded silent face and the conformationally flexible inner domain/bridging sheet, with its projecting variable loops, the identified site is recessed, although it must be accessible to a molecule as large as CD4 to serve its function. Otherwise, its size and reactive loops make it suitable for recognition by antibody. Recent serological analysis of long-term non-progressors attributes the broad and potent neutralizing properties of some of these sera to antibodies directed against this site[36]. The potential of receptor-binding-site antibodies to effect broad neutralization has been recognized since the first viral structures revealed conserved sites of receptor binding[37,38]. Here we show how an unanticipated requirement for the maintenance of a substantial receptor on-rate allows b12 to access and disable this most difficult of targets, the highly protected HIV-1 Env.

METHODS

Structure-based stabilization of gp120 in its CD4-bound conformation. Cavity-altering substitutions were identified by examination of highly conserved residues that bordered interdomain cavities, whereas all Cβ−Cβ distances between 3–7 Å were examined to identify potential stabilizing disulphides. A high-throughput expression system, coupling transient expression with swainsonine, endoglycosidase H and concanavalin A, was devised from related procedures[39] to produce gp120 suitable for crystallization. Briefly, codon-optimized variants were constructed in the context of core gp120 (HXBc2) and expressed transiently in 293 cells under control of the CMV/R promoter[40]. Swainsonine (20 mg l^{-1}; Biomol) was added to the cell-culture medium 2 h before DNA transfection, and supernatants were collected after 48 and 96 h. Secreted gp120 molecules were purified by 17b-affinity chromatography and deglycosylated with endoglycosidase H (Endo Hf; New England Biolab) at pH 5.9, 37 °C, with gp120 containing residual uncleaved glycan moieties removed by passage over concanavalin A Sepharose (Sigma-Aldrich). Ternary complexes of deglycosy-

lated core gp120, d1d2 of CD4 and 17b Fab were prepared as described previously[41], with $P222_1$ crystals grown (Supplementary Table 1) around conditions identified for the *Drosophila*-produced wild-type core[18]. Because crystals were small (rods, typical diameter 40 μm) and subject to high radiation dosages during data collection, difference Fourier maps comparing the initial and final swatches of data were inspected to identify radiation-induced disulphide breakage[42], and the refined models (Supplementary Table 1) adjusted to reflect the initial, radiation-damage-free structure.

Surface-plasmon resonance. A Biacore 3000 surface-plasmon resonance spectrometer was used to measure kinetic constants (see Supplementary Table 4 and Supplementary Fig. 5 for details).

Isothermal titration calorimetry (ITC). Wild-type core and variant gp120 molecules were dialysed in PBS buffer and titrated with d1d2 of CD4 on a MicroCal VP-ITC at 37 °C. Resultant data were examined using Origin software (MicroCal).

Structural determination of a b12–gp120 complex. The antigen-binding fragment (Fab) of b12 was produced by papain digestion, and purified with Superdex S200 chromatography (0.35 M NaCl, 2.5 mM Tris pH 7.1, 0.02% NaN_3). The Fab peak was pooled and mixed with deglycosylated gp120, which was produced by transient transfection as described above, and the resultant complexes purified by S200 chromatography. Crystals of Fab b12 and a two-disulphide variant (Ds12 F123) were grown by mixing 0.5 μl of complex (4 mg ml^{-1}) in S200 buffer with 0.5 μl of droplet mix (10.5% PEG 8,000, 0.2 M glycine, 105 mM Mg-acetate, 52.5 mM Na-cacodylate pH 6.5) and equilibrating in hanging droplets over reservoirs (droplet mix without glycine) at 20 °C. Hexagonal bi-pyramids (200 μm in length by 90 μm in diameter) were crosslinked[43], transferred to 15% PEG 8,000, 150 mM Mg-acetate, 100 mM Na-cacodylate pH 6.5, 30% ethylene glycol, 2.5% 2*R*,3*R*-butandiol, 2.5% trehalose, and flash-frozen in a nitrogen-cryostat stream. Data were collected at 100 K and processed with HKL2000 (ref. 44). Molecular replacement (AMoRe[45]) identified a 5.2σ peak (15–3 Å data) for the Fab portion (chains H and L) of the b12 IgG (Protein Data Bank 1HZH)[29], and phases from the rigid-body refined molecule allowed unambiguous placement of the outer domain in $F_o - F_c$ density. Iterative model building (XtalView[46]), combined with refinement (CNS[47], Refmac[48]), were used to define the remaining ordered parts of gp120.

Received 2 November 2006; accepted 8 January 2007.

1. Korber, B. *et al.* Timing the ancestor of the HIV-1 pandemic strain. *Science* **288,** 1789–1796 (2000).
2. Joint United National Programme on HIV/AIDS. 2006 Report on the global AIDS epidemic. ⟨http://www.aids.org/en/HIV data/2006GlobalReport/⟩ (2006).
3. Weiss, R. A. *et al.* Neutralization of human T-lymphotropic virus type III by sera of AIDS and AIDS-risk patients. *Nature* **316,** 69–72 (1985).
4. Wyatt, R. & Sodroski, J. The HIV-1 envelope glycoproteins: fusogens, antigens and immunogens. *Science* **280,** 1884–1888 (1998).
5. Parren, P. W., Moore, J. P., Burton, D. R. & Sattentua, Q. J. The neutralizing antibody response to HIV-1: viral evasion and escape from humoral immunity. *AIDS* **13** (suppl. A), S137–S162 (1999).
6. Kowalski, M. L. *et al.* Functional regions of the envelope glycoprotein of human immunodeficiency virus type 1. *Science* **237,** 1351–1355 (1987).
7. Lu, M., Blackow, S. & Kim, P. A trimeric structural domain of the HIV-1 transmembrane glycoprotein. *Nature Struct. Biol.* **2,** 1075–1082 (1995).
8. Starcich, B. R. *et al.* Identification and characterization of conserved and variable regions of the envelope gene HTLV-III/LAV, the retrovirus of AIDS. *Cell* **45,** 637–648 (1986).
9. Wyatt, R. *et al.* The antigenic structure of the human immunodeficiency virus gp120 envelope glycoprotein. *Nature* **393,** 705–711 (1998).
10. Kwong, P. D. *et al.* HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature* **420,** 678–682 (2002).
11. Wei, X. *et al.* Antibody neutralization and escape by HIV-1. *Nature* **422,** 307–312 (2003).
12. Burton, D. R. Antibodies, viruses and vaccines. *Nature Rev. Immunol.* **2,** 706–713 (2002).
13. Luftig, M. A. *et al.* Structural basis for HIV-1 neutralization by a gp41 fusion intermediate-directed antibody. *Nature Struct. Mol. Biol.* **13,** 740–747 (2006).
14. Dalgleish, A. G. *et al.* The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* **312,** 763–767 (1984).
15. Feng, F., Broder, C. C., Kennedy, P. E. & Berger, E. A. HIV-1 entry co-factor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272,** 872–877 (1996).
16. Chan, D. C., Fass, D., Berger, J. M. & Kim, P. S. Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **89,** 263–273 (1997).
17. Weissenhorn, W., Dessen, A., Harrison, S. C., Skehel, J. J. & Wiley, D. C. Atomic structure of the ectodomain from HIV-1 gp41. *Nature* **387,** 426–430 (1997).
18. Kwong, P. D. *et al.* Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393,** 648–659 (1998).
19. Thali, M. *et al.* Characterization of conserved human immunodeficiency virus type 1 (HIV-1) gp120 neutralization epitopes exposed upon gp120–CD4 binding. *J. Virol.* **67,** 3978–3988 (1993).
20. Chen, B. *et al.* Structure of an unliganded simian immunodeficiency virus gp120 core. *Nature* **433,** 834–841 (2005).
21. Zhang, M. Y. *et al.* Improved breadth and potency of an HIV-1-neutralizing single-chain antibody by random mutagenesis and sequential antigen panning. *J. Mol. Biol.* **335,** 209–219 (2004).
22. Rits-Volloch, S., Frey, G., Harrison, S. C. & Chen, B. Restraining the conformation of HIV-1 gp120 by removing a flexible loop. *EMBO J.* **25,** 5026–5035 (2006).
23. Profy, A. T. *et al.* Epitopes recognized by the neutralizing antibodies of an HIV-1-infected individual. *J. Immunol.* **144,** 4641–4647 (1990).
24. Burton, D. R. & Montefiori, D. C. The antibody response in HIV-1 infection. *AIDS* **11** (suppl. A), 587–598 (1997).
25. Burton, D. R. *et al.* Efficient neutralization of primary isolates of HIV-1 by a recombinant human monoclonal antibody. *Science* **266,** 1024–1027 (1994).
26. Trkola, A. *et al.* Cross-clade neutralization of primary isolates of human immunodeficiency virus type 1 by human monoclonal antibodies and tetrameric CD4 IgG. *J. Virol.* **69,** 6609–6617 (1995).
27. Trkola, A. *et al.* Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. *J. Virol.* **70,** 1100–1108 (1996).
28. Parren, P. W. *et al.* Antibody protects macaques against vaginal challenge with a pathogenic R5 simian/human immunodeficiency virus at serum levels giving complete neutralization *in vitro*. *J. Virol.* **75,** 8340–8347 (2001).
29. Saphire, E. O. *et al.* Crystal structure of a neutralizing human IgG against HIV-1: a template for vaccine design. *Science* **293,** 1155–1159 (2001).
30. Sundberg, E. J. & Mariuzza, R. A. Molecular recognition in antibody-antigen complexes. *Adv. Protein Chem.* **61,** 119–160 (2002).
31. Connolly, M. L. The molecular surface package. *J. Mol. Graph.* **11,** 139–141 (1993).
32. Pantophlet, R. *et al.* Fine mapping of the interaction of neutralizing and nonneutralizing monoclonal antibodies with the CD4 binding site of human immunodeficiency virus type 1 gp120. *J. Virol.* **77,** 642–658 (2003).
33. Zwick, M. B. *et al.* Molecular features of the broadly neutralizing immunoglobulin G1 b12 required for recognition of human immunodeficiency virus type 1 gp120. *J. Virol.* **77,** 5863–5876 (2003).
34. Yang, X. *et al.* Characterization of the outer domain of the gp120 glycoprotein from human immunodeficiency virus type 1. *J. Virol.* **78,** 12975–12986 (2004).
35. Arthos, J. *et al.* Biochemical and biological characterization of a dodecameric CD4-Ig fusion protein. Implications for therapeutic and vaccine strategies. *J. Biol. Chem.* **277,** 11456–11464 (2002).
36. Li, Y. *et al. Neutralizing specificity mapping in complex polyclonal sera* (AIDS Vaccine 2006, Amsterdam, 2006).
37. Rossmann, M. G. *et al.* Structure of a human common cold virus and functional relationship to other picornoviruses. *Nature* **317,** 145–153 (1985).
38. Wiley, D. C., Wilson, I. A. & Skehel, J. J. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. *Nature* **289,** 373–378 (1981).
39. Butters, T. D. *et al.* Effects of N-butyldeoxynojirimycin and the Lec3.2.8.1 mutant phenotype on N-glycan processing in Chinese hamster ovary cells: application to glycoprotein crystallization. *Protein Sci.* **8,** 1696–1701 (1999).
40. Barouch, D. H. *et al.* A human T-cell leukemia virus type 1 regulatory element enhances the immunogenicity of human immunodeficiency virus type 1 DNA vaccines in mice and nonhuman primates. *J. Virol.* **79,** 8828–8834 (2005).
41. Kwong, P. D. *et al.* Probability analysis of variational crystallization and its application to gp120, the exterior envelope glycoprotein of type 1 human immunodeficiency virus (HIV-1). *J. Biol. Chem.* **274,** 4115–4123 (1999).
42. Weik, M. *et al.* Specific chemical and structural damage to proteins produced by synchrotron radiation. *Proc. Natl Acad. Sci. USA* **97,** 623–628 (2000).
43. Lusty, C. J. A gentle vapor diffusion technique for cross-linking of protein crystals for cryocrystallography. *J. Appl. Crystallogr.* **32,** 106–112 (1999).
44. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
45. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50,** 157–163 (1994).
46. McRee, D. E. XtalView/Xfit—A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125,** 156–165 (1999).
47. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
48. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
49. Prabakaran, P. *et al.* Structure of severe acute respiratory syndrome coronavirus receptor-binding domain complexed with neutralizing antibody. *J. Biol. Chem.* **281,** 15829–15836 (2006).
50. Fleury, D., Daniels, R. S., Skehel, J. J., Knossow, M. & Bizebard, T. Structural evidence for recognition of a single epitope by two distinct antibodies. *Proteins* **40,** 572–578 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Graham, D. Hamer, S. Harrison, R. Pantophlet, W. Schief, L. Shapiro, I. Wilson and members of the Structural Biology Section,

VRC, for discussions and comments on the manuscript; M. Gao for assistance with PDB deposit; H. Katinger for antibody 2G12; G. Lin for suggesting the use of swainsonine; S. Majeed for preparation of Fab 17b; J. Nelson for assistance with b12 ELISAs; M. Posner for antibody F105; J. Robinson for antibodies 17b, 1.5e and F91; J. Stuckey for assistance with figures; M. Venturi for assistance with gp120 production methodology; and the NIH AIDS Research and Reference Reagent Program for CD4. Support for this work was provided by the Intramural Research Program of the NIH, by the International AIDS Vaccine Initiative, by a grant from the Bill and Melinda Gates Foundation Grand Challenges in Global Heath Initiative, and by grants from the NIH. Use of SER-CAT at the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Science.

Author Contributions T.Z. and P.D.K. carried out structure-based stabilization, SPR analyses and structural determinations; L.X. and G.J.N. constructed gp120 substitutions and developed and implemented a high-throughput gp120-production system suitable for crystallization; B.D. and R.W. carried out ITC characterizations; A.J.H., M.B.Z. and D.R.B. provided b12, b3, b6, b11 and b13, and mutant b12 binding; D.V.R. and J.A. provided D1D2-Igαtp and associated SPR analyses; S.-H.X., X.Y. and J.S. provided OD1 and preliminary design and antigenic analyses; and M.-Y.Z. and D.S.D. provided m6, m14 and m18. All authors contributed to the manuscript preparation.

Author Information Coordinates and structure factors have been deposited in the Protein Data Bank and may be obtained from the authors (accession codes 2nxy–2ny6 for the nine variant gp120 molecules with CD4 and 17b; accession code 2ny7 for the b12–gp120 complex). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.D.K. (pdkwong@nih.gov).

LETTERS

Early gas stripping as the origin of the darkest galaxies in the Universe

L. Mayer[1,2], S. Kazantzidis[3,4], C. Mastropietro[5] & J. Wadsley[6]

The known galaxies most dominated by dark matter (Draco, Ursa Minor and Andromeda IX) are satellites of the Milky Way and the Andromeda galaxies[1–4]. They are members of a class of faint galaxies, devoid of gas, known as dwarf spheroidals[3–5], and have by far the highest ratio of dark to luminous matter[3,6]. None of the models proposed to unravel their origin[7–10] can simultaneously explain their exceptional dark matter content and their proximity to a much larger galaxy. Here we report simulations showing that the progenitors of these galaxies were probably gas-dominated dwarf galaxies that became satellites of a larger galaxy earlier than the other dwarf spheroidals. We find that a combination of tidal shocks and ram pressure swept away the entire gas content of such progenitors about ten billion years ago because heating by the cosmic ultraviolet background kept the gas loosely bound: a tiny stellar component embedded in a relatively massive dark halo survived until today. All luminous galaxies should be surrounded by a few extremely dark-matter-dominated dwarf spheroidal satellites, and these should have the shortest orbital periods among dwarf spheroidals because they were accreted early.

Draco, Ursa Minor and Andromeda IX have mass-to-light ratios (M/L) larger than 100, but the majority of the other dwarf spheroidals in the Local Group have a lower M/L, of order 10–30, (ref. 1) typical among dwarf galaxies[11,12]. Another important difference is that Draco and Ursa Minor nearly stopped forming stars more than ten billion years ago, while other dwarf spheroidals continued to form stars for many billions of years[2]. The modest potential well of these extreme dwarfs cannot be the single property that determined their nature. Their halo masses are too large to invoke suppression of gas accretion owing to the cosmic ultraviolet background at high redshift[9,13] or blow-out due to supernovae winds[14]. Tidal shocks occurring as a dwarf repeatedly approaches the primary galaxy can transform rotationally supported systems resembling dwarf irregular galaxies into systems dominated by random motions, similar to dwarf spheroidals[10]. This tidal stirring can explain why dwarf spheroidals are more clustered around the primary galaxies relative to dwarf irregular galaxies but it leaves a significant gas component inside the dwarf, so that star formation can continue for several billions of years instead of being truncated early. Ram pressure in a hot gaseous corona could strip their gas completely[7,15] but the limitations of simulations so far have not allowed for firm predictions. For instance, existing calculations keep the structure of the stars and halo fixed in time and neglect radiative cooling and heating of the gas[16,17].

These earlier studies have explored the effect of a single gas removal mechanism and are, at best, only qualitatively consistent with the current structure formation paradigm, a model with cold dark matter and a cosmological constant (ΛCDM). Recent attempts to study the evolution of dwarf satellites directly in cosmological simulations rely on semi-analytical methods to model the baryonic component rather than solving the fluid equations[18]. These models neglect ram pressure, and because stripping by tides is slow and inefficient, they cannot explain the complete absence of gas and early truncation of star formation of the darkest dwarf spheroidals.

In cold-dark-matter models the present-day spatial distribution of subhalos within primary halos retains some indication of their infall time[19]. Satellites orbiting closer to the primaries were on average accreted earlier than those orbiting at larger distances. Interestingly, Draco and Ursa Minor lie at 68 and 86 kpc , respectively, from the Milky Way, and Andromeda IX at 45 kpc from Andromeda, while other dwarf spheroidals orbit as far as 200 kpc from the primaries[1]. We use a high-resolution ΛCDM dark-matter-only cosmological simulation of the formation of a Milky-Way-sized halo[20] (see also Supplementary Information). At redshift $z = 0$ we identify three subhaloes having distances less than 100 kpc from the centre and with peak circular velocities in the range 25–30 km s^{-1} (ref. 6) (Fig. 1). We track the orbits of the satellites back in time and find that two of them were accreted early, between $z = 2.5$ and $z = 1.5$.

We then construct a high-resolution n-body + smoothed particle hydrodynamics model of a dwarf galaxy satellite having a disk of stars and gas inside a cold-dark-matter halo (Fig. 1) with a peak velocity of about 40 km s^{-1}, comparable to that of the two identified cosmological subhalos before they were accreted onto the Milky Way (see Supplementary Information). We assume that 80% of the baryonic disk mass is in a gas component. This inefficient conversion of gas into stars is expected at these low-mass scales because most of the gas will have densities below the threshold for star formation[21]. In addition, at $z > 2$ the gas in the dwarf is heated to a temperature of over 10^4 K and ionized by the cosmic ultraviolet radiation, which further suppresses star formation.

The dwarf model is placed on an eccentric orbit inside a massive Milky-Way-sized halo model which is a replica of that in the cosmological simulation. We include radiative cooling as well as the heating and ionizing flux from the cosmic ultraviolet background radiation[22], and we embed a diffuse gaseous halo inside the dark halo of the primary. Such a halo is expected as a by-product of the process of galaxy formation and has a density and temperature consistent with observational constraints (see Supplementary Information).

With an orbital time of about 1.7 Gyr, the dwarf undergoes as many as five pericentre passages in 10 Gyr. At the first pericentre passage its dark halo loses 60% of its mass. The disk, deep inside the potential well of the halo, suffers no stripping, but the tidal perturbation triggers a strong bar instability (Fig. 1) and simultaneously heats the stars in the disk. The bar funnels most of the gas towards the central kiloparsec. Gas will be removed from the galaxy if the ram pressure force exerted by the diffuse hot gas, which is proportional to $\rho_g V^2$ (ρ_g being the density of the gaseous halo and V being the orbital speed of the galaxy), exceeds the gravitational restoring force provided by the potential well of the galaxy[23]. Ram pressure readily removes the gas outside the bar

[1]Institute for Theoretical Physics, University of Zurich, Winterthurestrasse 190, CH-8057 Zürich, Switzerland. [2]Institut für Astronomie, Department of Physics, ETH Zürich, Wolfgang-Pauli-Strasse 16, CH-8093 Zürich, Switzerland. [3]Kavli Institute for Particle Astrophysics and Cosmology, Department of Physics, Stanford University, MS 29, Stanford, California 94309, USA. [4]Kavli Institute for Cosmological Physics, Department of Astronomy and Astrophysics, The University of Chicago, Illinois 60637, USA. [5]Universitäts Sternwarte München, Scheinerstrasse 1, D-81679 München, Germany. [6]Department of Physics and Astronomy, McMaster University, Hamilton, Ontario L8S 4M1, Canada.

radius on a timescale of less than 10^8 years, but not the more tightly bound gas inside the bar. When the satellite crosses the pericentre a second time, the new tidal shock lowers the halo density by a factor of two inside a kiloparsec, so that V_{peak} drops to less than 30 km s^{-1}. The potential well has become shallower, so even the gas sitting inside the bar can be swept away by ram pressure (Fig. 1, see also Supplementary Information). The cosmic ultraviolet background heats and ionizes the gas, which enhances stripping significantly because the higher gas pressure opposes the gravitational restoring force. Once the first two orbits have been completed, no gas is retained by the dwarf.

The response of the system to the tides becomes progressively more impulsive at each new pericentre passage. As a result, the initially disk-like stellar distribution is heated into a nearly spherical, isotropic configuration (Fig. 1). The expansion reflects the attempt of the system to gain a new equilibrium as the internal binding energy is lowered. The removal of gas due to ram pressure is crucial for tidal heating to be effective (see Supplementary Information).

After ten billion years a diffuse spheroidal galaxy has replaced the gas-rich disk. During the last few orbits the stellar velocity dispersion profile is fairly flat (Fig. 2a), matching the profiles observed for Draco and Ursa Minor[24]. The initial angular momentum of the stars has been transported outward during the morphological transformation[8], producing a ratio between rotational velocity and the velocity dispersion (that is, the random motions) of $v_{rot}/\sigma < 0.2$ in the inner few hundred parsecs. The surface brightness and total luminosity resemble those of the faintest dwarf spheroidals (see legend of Fig. 2a). A substantial halo is preserved within a few kiloparsecs from the centre so that the final M/L is larger than 100 (Fig. 2b). The central dark matter density, which has decreased by almost a factor of four since the beginning of the simulation, is $\sim$0.2 $M_\odot$ pc^{-3}, comparable to that of Draco and Ursa Minor[1].

We predict that all massive galaxies should have a few extremely dark-matter-dominated satellites as the mechanism reported here is completely general within hierarchical structure formation. The efficiency of star formation in isolated low-mass galaxies can be lower[21] than that assumed here (see Supplementary Information), and hence dwarf satellites having gas fractions higher than 80% of the baryons before stripping probably did exist. Once they lose their gas, these systems will turn into dwarf spheroidals even more dark-matter-dominated than Draco and Ursa Minor. They should orbit the Milky Way and M31 and have escaped detection so far owing to their low surface brightness, especially at low Galactic latitude[25]. Their halo masses should be comparable to those of known dwarf spheroidals and thus might help to solve the 'missing satellites problem'[26]. The Ursa Major dwarf recently detected by the Sloan Digital Sky survey[27,28], which has a halo as massive as that of Draco but a luminosity a hundred times smaller, is probably one of them. Three more systems such as Ursa Major would be enough to bring theory and observations into agreement at the high end of the mass function of satellites, corresponding to a peak circular velocity of around 25–30 km s^{-1}, where suppression of baryonic infall by reionization is hardly effective.

When the ultraviolet background is an order of magnitude weaker, as is predicted[22] at $z < 1$, about 30% of the original gas component

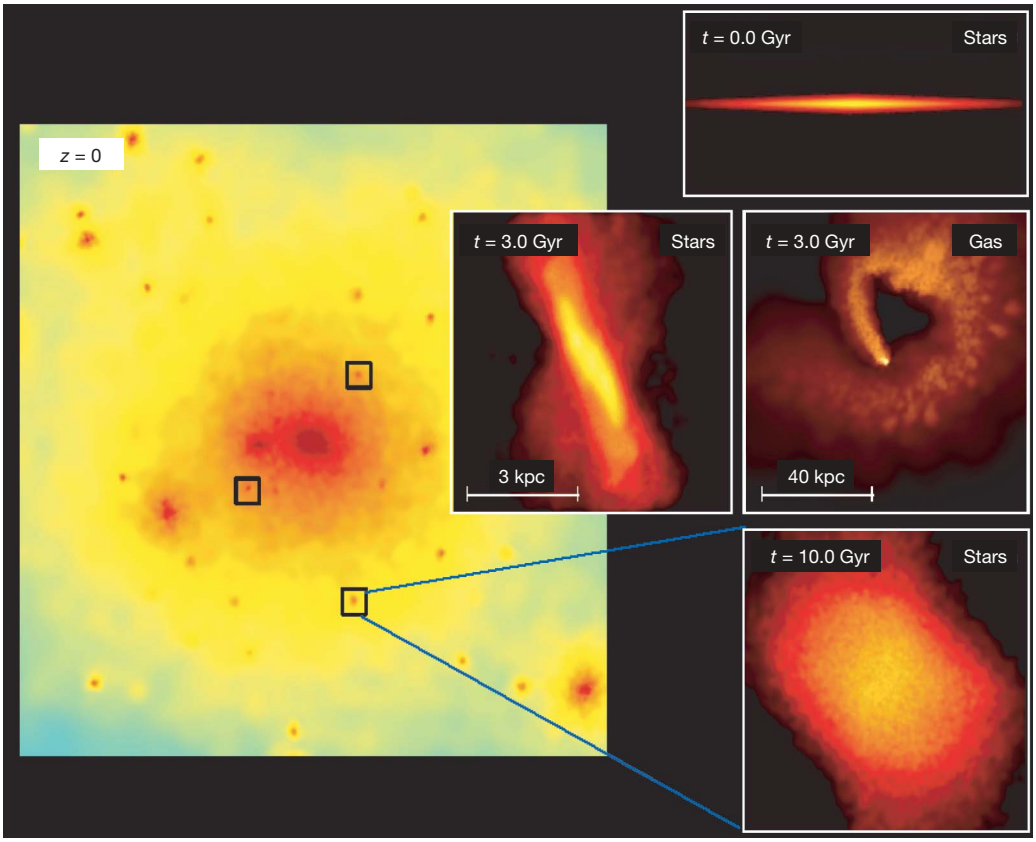


Figure 1 | Morphological evolution of the dwarf galaxy satellite. Colour-coded projected density map of the cosmological run at $z = 0$; the box is 260 kpc on a side, which corresponds to the virial radius of the Milky-Way-sized halo. The peak density along the line of sight is shown, ranging from 10^{-29}–10^{-24} g cm^{-3}, with the colour coding from blue (lowest density), through yellow, then red, to brown (highest density). The three satellites that meet the distance and circular velocity constraints (see text) are highlighted with black boxes. From top to bottom the insets show the stellar component of the dwarf galaxy, colour coded in projected density, at different times. Only regions with densities in the range 10^{-28}–10^{-23} g cm^{-3} are shown, with the colour coding from dark red (lowest density) to yellow (highest densities). In the top inset, the initial disk is shown edge-on (the box is 8 kpc on a side). In the left middle inset, the system is close to the second pericentre passage; the stars have assumed a strong bar-like configuration and heating is evident in the outskirts (the box is 7 kpc on a side). In the bottom inset, the end state is shown; the bar has been heated into a diffuse spheroid and any disk-like feature has been erased (the box is 4 kpc on a side and a projection along a random line of sight is shown). In the right middle inset, the trail of gas produced by ram pressure is shown, while even the residual gas in the centre is stripped. The colour-coded gas density projected onto the orbital plane is shown (densities in the range 10^{-30}–10^{-23} g cm^{-3}, colour coding as above), and the box is 100 kpc on a side.

stays in the galaxy (see Supplementary Information). Therefore dwarfs that fell into the Milky Way halo late can continue forming stars, and tidal shocks will produce periodic bursts of star formation[10]. These newcomers should account for those dwarf spheroidals that have fairly normal mass-to-light ratios, extended star formation histories and larger distances from the primaries[1,2]. This explains why Fornax is ten times brighter than Draco and has a very different star formation history despite having a comparable depth of potential well[1,6,29]. It implies that there should be a positive correlation between M/L and the infall epoch of dwarfs, and thus a negative correlation between M/L and their orbital time. The dwarf spheroidal Tucana represents the biggest challenge to our model because it lies far from any massive galaxy (see Supplementary Information). The accurate determination of the orbits of the dwarfs expected from ongoing and future astrometric missions such as the Space Inteferometry Mission and the Global Astrometric Interferometer for Astrophysics will be able to test this prediction.

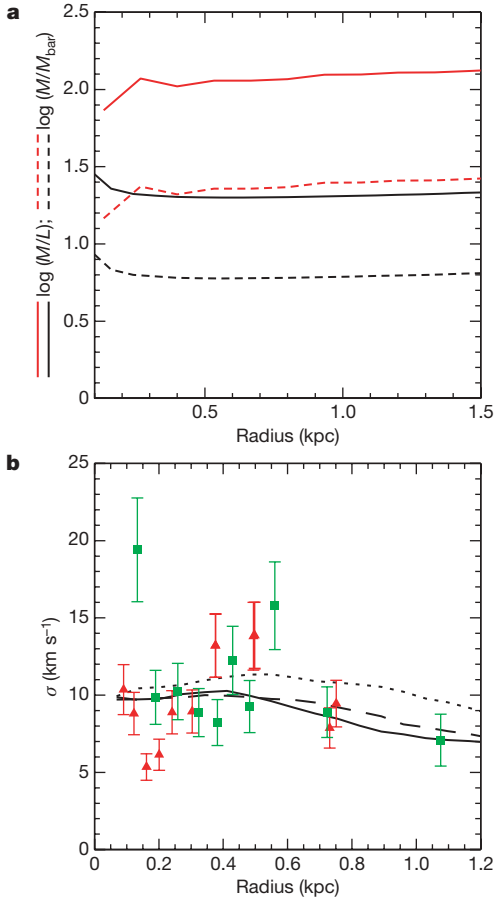


Figure 2 | Structural properties of the simulated dwarf after 10 billion years of evolution. a, Mass profiles shown out to the radius (from the centre of the simulated dwarf galaxy) at which stars are gravitationally bound. The dashed lines show the initial (black dashed line) and final (red dashed line) ratios of the total mass to the baryonic mass. The solid lines show the initial (black solid line) and final (red solid line) B band mass-to-light ratios of the dwarf. We have assumed stellar mass-to-light ratios of $M/L_{B^*} = 1.5$ (initial) and $M/L_{B^*} = 5$ (final) to compute the initial and final luminosities (as B band absolute magnitudes), $M_B = -12.5$ and $M_B = -9$. The final central surface brightness is $\mu_B \approx 26$ mag arcsec^{-2}. We note that using an initial $M/L_{B^*} = 1.5$ is motivated by the fact that at $z > 2$ a stellar population is at most three billion years old, whereas a final $M/L_{B^*} = 5$ is consistent with passive fading of the stellar population for about ten billion years[10]. **b**, The line-of-sight stellar velocity dispersion profiles are shown for three random directions (black lines) perpendicular to each other, together with published observational data points for Draco (red triangles) and Ursa Minor (green squares), including formal 1σ error bars[24]. The curves are shown out to the radius for which data points are available.

Received 11 July; accepted 22 December 2006.

1. Mateo, M. Dwarf galaxies of the local group. *Annu. Rev. Astron. Astrophys.* **36,** 435–506 (1998).
2. Grebel, E. K. in *The Stellar Content of Local Group Galaxies* (eds Whitelock, P. & Cannon, R.) 17–38 (IAU Symposium 192, ASP, San Francisco, 1999).
3. Kleyna, J. T., Wilkinson, M. I., Evans, N. W., Gilmore, G. & Frayn, C. Dark matter in dwarf spheroidals—II. Observations and modeling of Draco. *Mon. Not. R. Astron. Soc.* **330,** 778–791 (2002).
4. Chapman, S. *et al.* Keck DEIMOS kinematic study of Andromeda IX: Dark matter on the smallest galactic scales. *Astrophys. J.* **632,** L87–L90 (2005).
5. Gallagher, J. S., Madsen, G. J., Reynolds, R. J., Grebel, E. K. & Smecker-Hane, T. A. A search for ionized gas in the Draco and Ursa Minor dwarf spheroidal galaxies. *Astrophys. J.* **588,** 326–330 (2003).
6. Kazantzidis, S. *et al.* Density profiles of cold dark matter substructure: implications for the missing-satellites problem. *Astrophys. J.* **608,** 663–679 (2004).
7. Lin, D. N. C. & Faber, S. M. Some implications of nonluminous matter in dwarf spheroidal galaxies. *Astrophys. J.* **266,** L21–L25 (1983).
8. Dekel, A. & Silk, J. The origin of dwarf galaxies, cold dark matter, and biased galaxy formation. *Astrophys. J.* **303,** 39–55 (1986).
9. Bullock, J., Kravtsov, A. & Weinberg, D. H. Reionization and the abundance of galactic satellites. *Astrophys. J.* **539,** 517–521 (2000).
10. Mayer, L. *et al.* The metamorphosis of tidally stirred dwarf galaxies. *Astrophys. J.* **559,** 754–784 (2001).
11. de Blok, W. J. G. & McGaugh, S. S. The dark and visible matter content of low surface brightness disc galaxies. *Mon. Not. R. Astron. Soc.* **290,** 533–552 (1997).
12. Mayer, L. & Moore, B. The baryonic mass-velocity relation: clues to feedback processes during structure formation and the cosmic baryon inventory. *Mon. Not. R. Astron. Soc.* **354,** 477–484 (2004).
13. Susa, H. & Umemura, M. Formation of dwarf galaxies during the cosmic reionization. *Astrophys. J.* **600,** 1–16 (2004).
14. Mac Low, M. M. & Ferrara, A. Starburst-driven mass loss from dwarf galaxies: efficiency and metal ejection. *Astrophys. J.* **513,** 142–155 (1999).
15. Einasto, J., Kaasik, A. & Saar, E. Dynamic evidence on massive coronas of galaxies. *Nature* **252,** 111–113 (1974).
16. Marcolini, A., Brighenti, F. & D'Ercole, A. Three-dimensional simulations of the interstellar medium in dwarf galaxies—I. Ram pressure stripping. *Mon. Not. R. Astron. Soc.* **345,** 1329–1339 (2003).
17. Mori, M. & Burkert, A. Gas stripping of dwarf galaxies in clusters of galaxies. *Astrophys. J.* **538,** 559–568 (2000).
18. Kravtsov, A. V., Gnedin, O. Y. & Klypin, A. A. The tumultuous lives of galactic dwarfs and the missing satellites problem. *Astrophys. J.* **609,** 482–497 (2004).
19. Diemand, J., Madau, P. & Moore, B. The distribution and kinematics of early high-sigma peaks in present-day haloes: implications for rare objects and old stellar populations. *Mon. Not. R. Astron. Soc.* **364,** 367–383 (2005).
20. Governato, F. *et al.* The formation of a realistic disk galaxy in lambda-dominated cosmologies. *Astrophys. J.* **607,** 688–696 (2004).
21. Verde, L., Oh, P. S. & Jimenez, R. The abundance of dark galaxies. *Mon. Not. R. Astron. Soc.* **336,** 541–549 (2002).
22. Haardt, F. & Madau, P. Radiative transfer in a clumpy universe. II. The ultraviolet extragalactic background. *Astrophys. J.* **461,** 20–37 (1996).
23. Gunn, J. E. & Gott, J. R. I. On the infall of matter into clusters of galaxies and some effects on their evolution. *Astrophys. J.* **176,** 1–19 (1972).
24. Munoz, R. P. *et al.* Exploring halo substructure with giant stars: the velocity dispersion profiles of the Ursa Minor and Draco Dwarf spheroidal galaxies at large angular separations. *Astrophys. J.* **631,** L137–L141 (2005).
25. Willman, B., Governato, F., Dalcanton, J. J., Reed, D. & Quinn, T. The observed and predicted spatial distribution of Milky Way satellites. *Mon. Not. R. Astron. Soc.* **353,** 639–646 (2004).
26. Moore, B. *et al.* Dark matter substructure within galactic halos. *Astrophys. J.* **524,** L19–L22 (1999).
27. Willman, B. *et al.* A new Milky Way dwarf galaxy in Ursa Major. *Astrophys. J.* **626,** L85–L88 (2005).
28. Kleyna, J. T., Wilkinson, M. I., Evans, N. W. & Gilmore, G. Ursa Major:a missing low mass CDM halo? *Astrophys. J.* **630,** L141–L144 (2005).
29. Walker, M. G. *et al.* Internal kinematics of the Fornax dwarf spheroidal galaxy. *Astron. J.* **131,** 2114–2139 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Y. Birinboim, A. Dekel, F. Governato, A. Kravtsov, G. Lake, C. Porciani, M. Valluri, B. Willman and A. Zentner, for discussions. We also thank S. Majewski and R. Munoz for sharing their data with us. L. Mayer and S. Kazantzidis are grateful to the Aspen Center for Physics, where some of this work was completed. All computations were performed on the Zbox supercomputer at the University of Zürich, on LeMieux at the Pittsburgh Supercomputing Center, and on the Gonzales cluster at ETH Zürich.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.M. (lucio@phys.ethz.ch) or S.K. (stelios@kicp.uchicago.edu).

LETTERS

Energy doubling of 42 GeV electrons in a metre-scale plasma wakefield accelerator

Ian Blumenfeld[1], Christopher E. Clayton[2], Franz-Josef Decker[1], Mark J. Hogan[1], Chengkun Huang[2], Rasmus Ischebeck[1], Richard Iverson[1], Chandrashekhar Joshi[2], Thomas Katsouleas[3], Neil Kirby[1], Wei Lu[2], Kenneth A. Marsh[2], Warren B. Mori[2], Patric Muggli[3], Erdem Oz[3], Robert H. Siemann[1], Dieter Walz[1] & Miaomiao Zhou[2]

The energy frontier of particle physics is several trillion electron volts, but colliders capable of reaching this regime (such as the Large Hadron Collider and the International Linear Collider) are costly and time-consuming to build; it is therefore important to explore new methods of accelerating particles to high energies. Plasma-based accelerators are particularly attractive because they are capable of producing accelerating fields that are orders of magnitude larger than those used in conventional colliders[1–3]. In these accelerators, a drive beam (either laser or particle) produces a plasma wave (wakefield) that accelerates charged particles[4–11]. The ultimate utility of plasma accelerators will depend on sustaining ultrahigh accelerating fields over a substantial length to achieve a significant energy gain. Here we show that an energy gain of more than 42 GeV is achieved in a plasma wakefield accelerator of 85 cm length, driven by a 42 GeV electron beam at the Stanford Linear Accelerator Center (SLAC). The results are in excellent agreement with the predictions of three-dimensional particle-in-cell simulations. Most of the beam electrons lose energy to the plasma wave, but some electrons in the back of the same beam pulse are accelerated with a field of ~52 GV m^{-1}. This effectively doubles their energy, producing the energy gain of the 3-km-long SLAC accelerator in less than a metre for a small fraction of the electrons in the injected bunch. This is an important step towards demonstrating the viability of plasma accelerators for high-energy physics applications.

In a plasma wakefield accelerator large-amplitude electric fields result from space-charge waves excited by the passage of an ultra-relativistic electron beam through a plasma[12]. A fully ionized plasma can be formed in a neutral vapour when the radial electric field of the electron beam exceeds the field ionization threshold[13]. The ionization occurs in a narrow region in the front of the beam. This ionization front produces a plasma that has a radius much larger than the beam itself. If the beam density exceeds the plasma density, the plasma electrons are expelled from the volume of the electron pulse, leaving a column of more massive ions behind[14]. Subsequently, the expelled plasma electrons are pulled back (by the ions) to the beam axis behind the pulse, overshoot, and set up a space-charge oscillation or wake. The longitudinal field of this wake varies continuously along the pulse, decelerating its core but accelerating the particles in the back. The ion column also provides a focusing force[15] that guides the beam over many diffraction lengths, allowing an efficient transfer of the beam energy to the wake. This force also causes the transverse size of the beam to oscillate as it propagates through the plasma—the so-called betatron oscillations (see Supplementary Movie 1).

Recent plasma wakefield accelerator experiments have shown high-gradient acceleration of electrons using a 10-cm-long plasma[11]. To obtain energy gains of interest to high-energy physics, these high gradients must be extended over metre-scale plasmas. Such an extension transitions the plasma wakefield accelerator from a regime in which the drive beam has no time to distort, deplete or go unstable, to a regime in which it is significantly depleted in energy, deformed owing to combined effects of diffraction and multiple transverse oscillations, and possibly goes unstable because of the electron-hose instability[16]. This work is in this latter regime.

A schematic of the experimental set-up is shown in Fig. 1. In the present work carried out at the Final Focus Test Beam facility at SLAC, the nominally 50-femtosecond-long electron beam containing 1.8×10^{10} particles is focused to a spot size of ~10 μm at the entrance of an 85-cm-long column of lithium vapour with a density n_e of 2.7×10^{17} cm^{-3}. The nominally 42 GeV beam has a correlated energy spread of approximately 1.5 GeV, with electrons in the front of the beam at higher energies than those at the back. The beam exiting the plasma traverses a metre-long dipole magnet, which disperses the beam electrons according to their energy. The transverse distribution of the dispersed electrons is measured at two distances (planes 1 and 2 in Fig. 1) downstream of the dipole magnet to distinguish the energy changes of the electrons from their possible transverse deflection due to the plasma.

Images of the dispersed electrons are recorded along with the relevant beam parameters on a shot-to-shot basis. The energy gain achieved for each shot is determined as described in the Methods section. Figure 2 shows one example of the electron energy

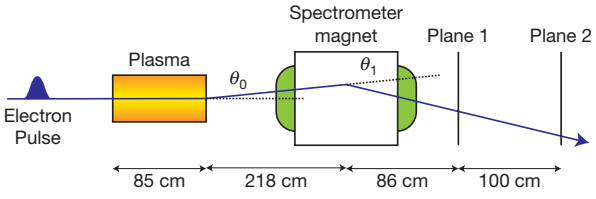


Figure 1 | Schematic of the experimental set-up. Two cameras record the energy-dispersed images at planes 1 and 2. A combination of low dispersion at plane 1 and a lower lens magnification on the camera allows a broad energy spectrum of the beam, including energy gain and loss, to be recorded. A higher dispersion at plane 2 coupled with a larger lens magnification is used to record images showing greater detail of the energy gain. The comparison of these two images allows for an independent measurement of vertical deflection and energy gain, as discussed in the Methods section.

[1]Stanford Linear Accelerator Center, 2575 Sand Hill Road, Menlo Park, California 94025, USA. [2]University of California Los Angeles, 405 Hilgard Avenue, Los Angeles, California 90095, USA. [3]University of Southern California, University Park, Los Angeles, California 90089, USA.

distribution between 35 and 100 GeV after traversing the plasma. The angle θ_0 at the plasma exit for this particular event was calculated to be smaller than 100 μrad, which is negligible; therefore energy relates directly to position. The highest electron energy is 85 ± 7 GeV, indicating that some electrons in the tail of the beam with an initial energy of 41 GeV have more than doubled their initial energy. The implied peak accelerating field of ~52 GV m^{-1} is consistent with the fields previously measured in a 10-cm-long plasma[11], indicating that the energy gain is scalable by extending the length of the plasma at least up to 85 cm. With this plasma length, in a series of 800 events, 30% showed an energy gain of more than 30 GeV. Variations in the measured energy gain were correlated to fluctuations in the peak current of the incoming electron beam.

When the length of the lithium vapour column was extended from 85 cm to 113 cm, the maximum energy in an event with a similar incoming current profile was measured to be 71 ± 11 GeV. Less than 3% of a sample of 800 consecutive events showed an energy gain of more than 30 GeV. There are three possible reasons for this apparent saturation of energy gain observed in the experiment. The first is that the energy of the particles that produced the wake has been depleted to almost zero, such that the acceleration is terminated in the last 28 cm of the plasma. However, the minimum energy measured at plane 1 (not shown) was 5–7 GeV, which is inconsistent with this explanation. The second possible reason is that the electron hosing instability is so severe that the beam breaks up[16]. In the data shown in Fig. 2 there are negligible transverse deflections of the various longitudinal slices of the beam, indicating an absence of the hosing instability. The third possibility is head erosion: the front of the beam expands, because it is not subjected to the focusing force of the ion column. This expansion decreases the beam density, which moves the ionization front backward in the beam frame. Eventually the beam electric field drops below the threshold for plasma formation, terminating the acceleration process before the energy of the drive beam is depleted (see Supplementary Movie 1).

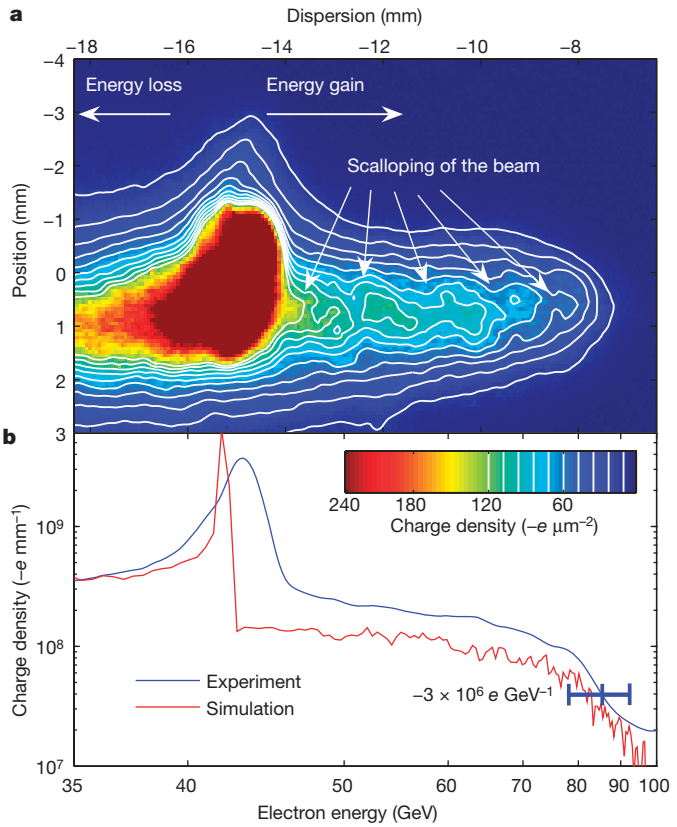


Figure 2 | Energy spectrum of the electrons. a, Energy spectrum of the electrons in the 35–100 GeV range as observed in plane 2. The dispersion (shown on the top axis) is inversely proportional to the particle energy (shown on the bottom axis). The head of the pulse, which is unaffected by the plasma, is at 43 GeV. The core of the pulse, which has lost energy driving the plasma wake, is dispersed partly out of the field of view of the camera. Particles in the back of the bunch, which have reached energies up to 85 GeV, are visible to the right. The pulse envelope exits the plasma with an energy-dependent betatron phase advance, which is consistent with the observed scalloping of the dispersed beam. **b**, Projection of the image in **a**, shown in blue. The simulated energy spectrum is shown in red. The differences between the measured and the simulated spectrum near 42 GeV are due to an initial correlated energy spread of 1.5 GeV not included in the simulations. The horizontal error bar is due to the uncertainty in estimating the deflection angle and the spot size of the beam.

We used simulations to explain the maximum electron energy observed in the experiment. Figure 2b shows a comparison of the measured energy spectrum with one derived from simulations. The electron current distribution is extracted from the energy spectrum of the beam measured upstream of the plasma by comparing it to a phase space simulation using the code LiTrack[17]. The wakefield from this current distribution and the propagation of the pulse through the plasma are modelled using the three-dimensional, parallel particle-in-cell (3D-PIC) code QuickPIC[18]. QuickPIC includes the effects of field ionization and electron energy loss due to radiation[19] from oscillations in the ion column.

Figure 3a and b shows the simulation output at two different positions in the plasma. At a distance of 12.3 cm, the wake produced by the motion of the plasma electrons resembles that produced in a preformed plasma, because the ionization occurs near the very head of the beam. The expelled plasma electrons return to the beam axis at nearly the same z location. This gives rise to an extremely large spike in the accelerating field. After 81.9 cm one can see the effect of beam head erosion in that the ionization front now occurs further back along the pulse. Even though the wake is formed further back, the peak accelerating field occurs at approximately the same position along the pulse. The transverse size of the pulse ahead of the ionization front is so large that the local beam density has dropped below the useful range in the colour table. However, the modified ionization front causes some blurring of the position at which the returning plasma electrons arrive on the axis, an effect known as phase mixing. This not only reduces the peak accelerating field but also leads to some defocusing of the high-energy beam electrons in this region (see Supplementary Discussion and Supplementary Figs 1–4).

The simulated energy distribution at this point was binned equivalently to the experimental data, as shown in Fig. 2b. The quantitative agreement between the two spectra is good. In the simulation spectrum of Fig. 2b, electrons are accelerated to a maximum energy of 95 GeV. In the experiment, the maximum detectable energy is determined by the spot size at the detection plane, and the highest detected energy is 85 GeV. For the present case, this corresponds to a detection threshold of 3×10^6 electrons per GeV. The mean electron energy of the highest energy bin containing 3×10^6 electrons per GeV in the simulation is shown as a function of position along the plasma in Fig. 3c. Also shown are maximum energies measured in the experiment at 85 and 113 cm for similar electron current profiles. The energy in the simulation increases approximately linearly with propagation distance up to a value of 80 GeV at about 70 cm and then saturates at 85 GeV at 85 cm owing to the phase-mixing effect, which leads to gradual defocusing of the highest energy electrons as mentioned above. As the beam propagates beyond 85 cm, the highest-energy electrons continue to be defocused to such an extent that at 104 cm a significant number of the high-energy electrons are lost to the simulation walls, causing the maximum observed electron energy to drop to 60 GeV. In the experiment, electrons defocused at such angles would not be detectable in the electron spectrometer. It should be noted that no significant wakefield is left beyond 104 cm, because the electron beam core containing the bulk of the particles is completely eroded away.

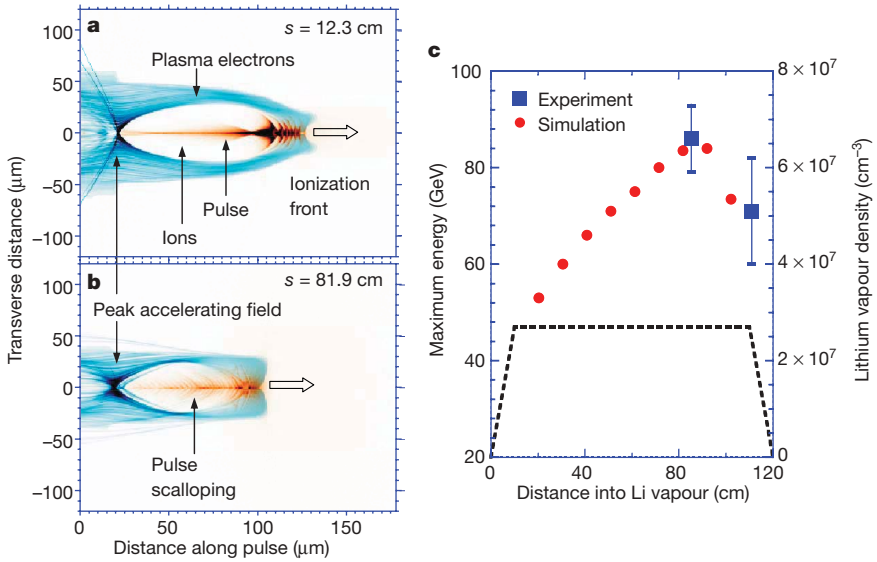


Figure 3 | Simulation of the experiment using the code QuickPIC. The density of the electron pulse (brown) and the plasma electrons (blue) at the distance the beam pulse has propagated $s = 12.3$ cm (**a**) and 81.9 cm (**b**) into the plasma on a plane ($y = 0$) through the centre of the simulation box. The pulse travels from left to right. The scalloping features seen at the front of the pulses in **a** and **b** are the result of an increasing focusing force as the plasma electrons are still being blown out by the beam electrons. The back of the pulse, entirely within the uniform ion column, is nearly uniformly focused in **a**. However, the scalloping of the back of the pulse in **b**—which now has a wide range of energies—is due to the energy-dependent focusing through the ion column. Similar features are identifiable in the experimental data of Fig. 2a. **c**, The maximum observed energy in the experiment (blue squares) for two different plasma lengths is compared to the energy of the particle bin containing 3×10^6 electrons GeV^{-1} (approximately the experimental detection threshold) in simulations (red dots) as a function of distance in the laboratory frame. Also shown is the lithium density profile used for the simulations (dashed line). Vertical error bars are due to the uncertainty in estimating the deflection angle and the spot size of the beam.

The simulations have reproduced the energy spectrum seen in the experiment for an 85-cm-long plasma and elucidated the underlying physical mechanism, head erosion, which leads to the observed saturation of the maximum energy. This effect can be avoided by the use of a lower-emittance beam such that its diffraction length is longer than the plasma length. In that case, the maximum energy gain would be determined by the energy of the drive beam.

Thus, we have produced an accelerating field of 52 GV m^{-1} in a plasma wakefield accelerator and sustained it for 85 cm. The result is in excellent agreement with 3D-PIC simulations. By producing the energy gain of the 3-km-long SLAC accelerator in less than a metre, albeit for a relatively small number of electrons, we have taken an important step towards demonstrating the viability of plasma accelerators for high-energy physics.

METHODS

Electron pulses. A 6-mm-long electron pulse from the SLAC damping ring undergoes three stages of longitudinal compression. In each of these stages, a time-dependent energy is added to the pulse, which is followed by magnetic transport elements that compress the pulse. As a result, the originally 6-mm-long pulses are compressed by a factor of 500 to a minimum length of 12 μm (ref. 20). Such a large compression is sensitive to the phases and amplitudes of the klystrons powering the accelerating sections of the linear accelerator, leading to some pulse-to-pulse variation in the bunch length.

The beam has geometric transverse emittances of $\varepsilon_x = 9.5 \times 10^{-10}$ m and $\varepsilon_y = 1.2 \times 10^{-10}$ m. It is focused with a quadrupole doublet to a spot with 10 μm radius at the entrance of the plasma. With this beam energy, bunch length and spot size, the corresponding power density is 3×10^{20} W cm^{-2}.

Plasma generation. A column of lithium vapour with a density of 2.7×10^{17} cm^{-3} is produced in a heat-pipe oven[21]. The lithium vapour is confined by a helium buffer gas, which is in turn separated from the beam-line vacuum by a 50-μm-thick beryllium window upstream and by a 75-μm-thick beryllium window downstream. Lithium was chosen because of the low ionization potential of its first electron (5.4 eV) and the relatively high potential for its two subsequent electrons (76 and 122 eV). In the present experiments the transverse electric field of the ultrashort electron pulses is large enough to field-ionize the first lithium electron over a timescale shorter than the bunch duration. The ADK theory for field ionization[22] indicates that full ionization occurs in the volume surrounding the pulse in which the electric field exceeds ~6 GV m^{-1}. Thus, the full ionization extends over a radius of more than 100 μm and ionization begins far earlier than the peak of the bunch current. Because the ionization region extends over a radius larger than the plasma collisionless skin depth (c/ω_p, where $\omega_p = (n_e e^2/\varepsilon_0 m_e)^{1/2}$ is the plasma angular frequency; e is the charge on the electron, ε_0 is the permittivity of free space and m_e is the mass of the electron), the wake is similar to that in a preformed plasma.

Energy measurement. The energy spectrometer consists of a dipole magnet that disperses the electrons vertically according to their momentum p. The dispersion can be closely approximated by a deflection at the centre of the magnet: $\theta_1 = e\int B\mathrm{d}L/p$. Using the measured dispersion, its integrated magnetic flux density $\int B\mathrm{d}L$ was calculated to be 1.2 T m. In general, all particles in a pulse leave the plasma from a well-defined spot, but with a non-negligible exit angle θ_0. To discriminate between a vertical exit angle and the deflection by the magnet, the particle distribution is measured at two planes, 86 cm and 186 cm downstream of the centre of the dipole (Fig. 1).

At each of the two planes, the particle distribution is measured by imaging Cherenkov radiation emitted as the electrons pass through a 15-mm-wide air gap established by two silicon wafers (not shown in Fig. 1), positioned at an angle of 45° to the beam. The second wafer acts as a mirror and deflects the Cherenkov light into a lens that images the origin of the light onto a cooled charge-coupled device camera (CCD). The electrons pass the silicon almost unperturbed.

A system of equations is set up relating the offsets at the two planes to two angles, the exit angle at the plasma θ_0 and the deflection angle in the magnet θ_1 (see Fig. 1). For each feature in the spectrum that can be identified on both screens, for instance scalloping of the beam shown in Fig. 2a, this system of equations has been solved for θ_0 and θ_1, the latter angle giving the particle energy. The highest-energy feature that can clearly be resolved (see Fig. 2a) is used to determine the energy gain for this event. The uncertainty in the energy measurement is dominated by the uncertainty in the determination of the position of this feature.

The images have been corrected at the level of a few per cent for the non-uniform collection efficiency of the optics. Pixel-to-pixel variations in the CCD offset and a common mode have been subtracted; the signal from X-rays that hit the CCD directly has been eliminated.

Simulations. The simulations were done using the quasi-static, three-dimensional, particle-in-cell code called QuickPIC. The three-dimensional computational grid forms a box xyz (240 μm × 240 μm × 260 μm) in size whose axial coordinate is z-ct. Therefore, the simulation window moves at the speed of light, which is very close to the beam speed in the z direction. The number of grid points is 256 × 256 × 512, respectively. The beam is initialized so that in vacuum, it would focus 15 cm beyond the start of the lithium vapour with a 10 μm root-mean-square spot size. The

longitudinal current profile is extracted from the unique LiTrack simulation that matches the experimentally measured beam spectrum produced by the SLAC accelerator. The resulting current profile approximates a gaussian ($\sigma_z \approx 15\,\mu m$) with a small tail. We use 8.4 million particles for the beam and 2.6×10^5 particles for each 'slice' of lithium. In the quasi-static approximation, as the entire beam moves through a slice of gas, the lithium ionizes, the resulting plasma evolves transversely and, to account for the axial motion, the charge on each particle is suitably changed. The resulting plasma forces are stored for each slice and are then used to advance the momentum and position of each beam electron. The beam electrons are advanced every 1.0 mm, which is 1/26th of a betatron wavelength for 42 GeV electrons in the flat density region. The simulations were done on the Apple X-serve Dawson Cluster at UCLA.

Received 21 July; accepted 13 December 2006.

1. Tajima, T. & Dawson, J. M. Laser electron accelerator. *Phys. Rev. Lett.* **43,** 267–270 (1979).
2. Chen, P. *et al.* Acceleration of electrons by the interaction of a bunched electron beam with a plasma. *Phys. Rev. Lett.* **54,** 693–696 (1985).
3. Joshi, C. *et al.* Ultrahigh gradient particle acceleration by intense laser-driven plasma density waves. *Nature* **311,** 525–529 (1984).
4. Modena, A. *et al.* Electron acceleration from the breaking of relativistic plasma waves. *Nature* **377,** 606–608 (1995).
5. Gordon, D. *et al.* Observation of electron energies beyond the linear dephasing limit from a laser-excited relativistic plasma wave. *Phys. Rev. Lett.* **80,** 2133–2136 (1998).
6. Umstadter, D. *et al.* Nonlinear optics in relativistic plasmas and laser wake field acceleration of electrons. *Science* **273,** 472–475 (1996).
7. Mangles, S. P. D. *et al.* Monoenergetic beams of relativistic electrons from intense laser–plasma interactions. *Nature* **431,** 535–538 (2004).
8. Geddes, C. G. R. *et al.* High-quality electron beams from a laser wakefield accelerator using plasma-channel guiding. *Nature* **431,** 538–541 (2004).
9. Faure, J. *et al.* A laser–plasma accelerator producing monoenergetic electron beams. *Nature* **431,** 541–544 (2004).
10. Barov, N. *et al.* Propagation of short electron pulses in a plasma channel. *Phys. Rev. Lett.* **80,** 81–84 (1998).
11. Hogan, M. J. *et al.* Multi-GeV energy gain in a plasma-wakefield accelerator. *Phys. Rev. Lett.* **95,** 054802 (2005).
12. Katsouleas, T. Physical mechanisms in the plasma wake-field accelerator. *Phys. Rev. A.* **33,** 2056–2064 (1986).
13. Bruhwiler, D. *et al.* Particle-in-cell simulations of tunneling ionization effects in plasma-based accelerators. *Phys. Plasmas* **10,** 2022–2030 (2003).
14. Rosenzweig, J. B. *et al.* Acceleration and focusing of electrons in two-dimensional nonlinear plasma wake fields. *Phys. Rev. A.* **44,** R6189–R6192 (1991).
15. Clayton, C. E. *et al.* Transverse envelope dynamics of a 28.5-GeV electron beam in a long plasma. *Phys. Rev. Lett.* **88,** 154801 (2002).
16. Dodd, E. S. *et al.* Hosing and sloshing of short-pulse GeV-class wakefield drivers. *Phys. Rev. Lett.* **88,** 125001 (2002).
17. Bane, K. L. F. & Emma, P. LiTrack: a fast longitudinal phase space tracking code with graphical user interface. Stanford Linear Accelerator Center Report No. SLAC-PUB-11035 (SLAC, Menlo Park, California, 2005).
18. Huang, C. *et al.* QUICKPIC: A highly efficient particle-in-cell code for modeling wakefield acceleration in plasmas. *J. Comput. Phys.* **217,** 658–679 (2006).
19. Johnson, D. K. *et al.* Positron production by X rays emitted by betatron motion in a plasma wiggler. *Phys. Rev. Lett.* **97,** 175003 (2006).
20. Krejcik, P. *et al.* Commissioning of the SPPS linac bunch compressor. *Proceedings of the Particle Accelerator Conference (12–16 May, 2003, Portland, Oregon)* 423–425 (IEEE, Piscataway, New Jersey, 2003).
21. Muggli, P. *et al.* Photo-ionized lithium source for plasma accelerator applications. *IEEE Trans. Plasma Sci.* **27,** 791–799 (1999).
22. Ammosov, M. V. *et al.* Tunnel ionization of complex atoms and of atomic ions in an alternating electromagnetic field. *Sov. Phys. JETP* **64,** 1191–1194 (1986).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the US Department of Energy and the National Science Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.J. (joshi@ee.ucla.edu).

LETTERS

Ultralow-power organic complementary circuits

Hagen Klauk[1], Ute Zschieschang[1], Jens Pflaum[2] & Marcus Halik[3]

The prospect of using low-temperature processable organic semiconductors to implement transistors, circuits, displays and sensors on arbitrary substrates, such as glass or plastics, offers enormous potential for a wide range of electronic products[1]. Of particular interest are portable devices that can be powered by small batteries or by near-field radio-frequency coupling. The main problem with existing approaches is the large power consumption of conventional organic circuits, which makes battery-powered applications problematic, if not impossible. Here we demonstrate an organic circuit with very low power consumption that uses a self-assembled monolayer gate dielectric and two different air-stable molecular semiconductors (pentacene and hexadecafluorocopperphthalocyanine, F_{16}CuPc). The monolayer dielectric is grown on patterned metal gates at room temperature and is optimized to provide a large gate capacitance and low gate leakage currents. By combining low-voltage p-channel and n-channel organic thin-film transistors in a complementary circuit design, the static currents are reduced to below 100 pA per logic gate. We have fabricated complementary inverters, NAND gates, and ring oscillators that operate with supply voltages between 1.5 and 3 V and have a static power consumption of less than 1 nW per logic gate. These organic circuits are thus well suited for battery-powered systems such as portable display devices[2] and large-surface sensor networks[3] as well as for radio-frequency identification tags with extended operating range[4].

The gate–source voltage V_{GS} required to induce a charge density Q in the channel of a field-effect transistor is determined by the gate dielectric capacitance C: $V_{GS} = Q/C = Qt/\varepsilon\varepsilon_0$, where t and ε are the thickness and permittivity of the dielectric, respectively. Thus the operating voltage of a transistor scales with the dielectric thickness. A promising path to high-mobility, low-voltage organic transistors is the use of gate dielectrics based on molecular self-assembled monolayers (SAMs) that provide a capacitance approaching 1 μF cm^{-2} and thus allow organic transistors to operate with voltages of a few volts. So far, SAM-dielectric organic transistors have been demonstrated only on silicon[5–8] and on substrates coated with indium tin oxide[7]. The implementation of high-performance integrated circuits, however, requires a gate material with large electrical conductivity that can be deposited at low temperatures on an insulating substrate to define individual gate electrodes for each transistor. Here we describe how to grow high-quality SAM dielectrics with very small leakage currents on evaporated, patterned aluminium gates. We found that the charge leakage through SAMs on aluminium is no larger than the leakage through high-quality SAMs on silicon[5–7], suggesting that SAMs on aluminium can be densely packed, as in the case of SAMs on silicon.

Whereas the formation of SAMs on natively oxidized silicon is best facilitated with silane anchor groups[9], we have chosen phosphonic acid anchor groups for self-assembly on aluminium[10–12]. Figure 1a shows the chemical structure of n-octadecylphosphonic acid, which we selected for this work. SAMs were prepared in a solution of 2-propanol at room temperature[12]. Substrates were glass or flexible polyethylene naphthalate (for transistors, circuits and leakage-current test structures) or silicon (for X-ray measurements). Aluminium 20 nm thick was deposited by thermal evaporation. Before self-assembly the aluminium surface was briefly exposed to an oxygen plasma (150 W, 15 s). The plasma treatment increases the thickness of the (native) aluminium oxide layer and creates a density of hydroxyl groups sufficient for molecular adsorption. Small-angle

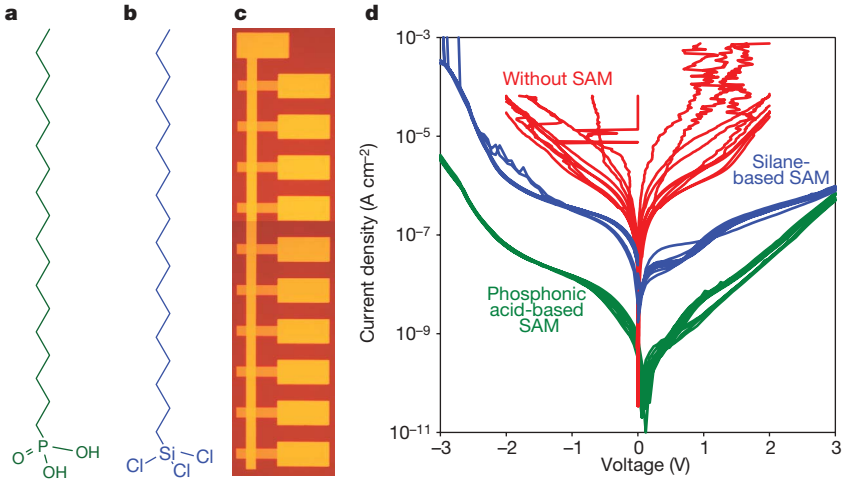


Figure 1 | Self-assembled monolayer dielectrics on metal gates. **a**, Chemical structure of n-octadecylphosphonic acid. **b**, Chemical structure of n-octadecyltrichlorosilane. **c**, Photograph of the leakage current test structure with shadow-mask-patterned aluminium bottom and gold top electrodes. **d**, Leakage current density as a function of applied voltage. Each measurement was repeated on ten junctions to evaluate the uniformity.

[1]Max Planck Institute for Solid State Research, Heisenbergstrasse 1, 70569 Stuttgart, Germany. [2]University Stuttgart, Third Institute of Physics, Pfaffenwaldring 57, 70550 Stuttgart, Germany. [3]Friedrich-Alexander University Erlangen-Nürnberg, Institute of Polymer Materials, Martensstrasse 7, 91058 Erlangen, Germany.

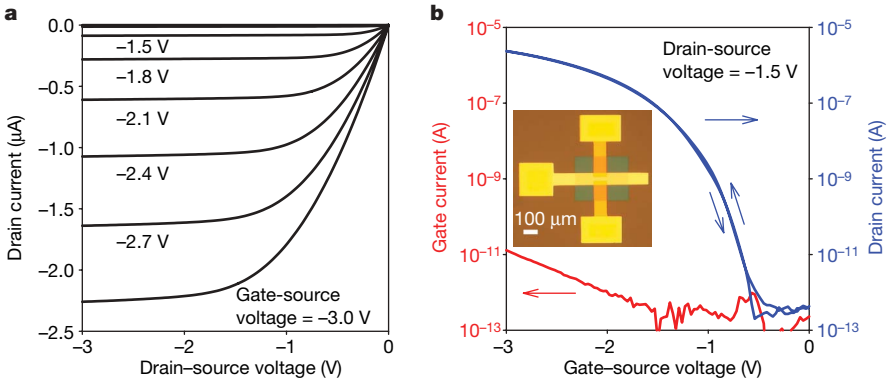


Figure 2 | p-channel pentacene TFT with SAM gate dielectric. The TFT has a channel length of 30 μm and a channel width of 100 μm. **a**, Drain current as a function of drain–source voltage. **b**, Drain current and gate current as a function of gate–source voltage (the inset is a photograph of the transistor). Owing to the small thickness and the good insulating properties of the dielectric the TFT can be operated with low voltages and has a maximum gate current of only 15 pA. The carrier mobility is 0.6 $cm^2 V^{-1} s^{-1}$.

X-ray reflectivity measurements indicate that the thickness of the aluminium oxide increases from 1.8 to 3.8 nm during the plasma treatment (see Supplementary Fig. 1). Static contact angles are 10° before and 105° after SAM preparation, confirming the formation of a hydrophobic monolayer[10]. From the reflectivity measurements we conclude that the SAM has a thickness of 2.1 nm. Assuming the molecules are 2.3 nm long (as calculated using CambridgeSoft Chem3D Pro; www.cambridgesoft.com), this suggests that the molecules stand approximately upright with a tilt angle of about 24° with respect to the surface normal.

Although the SAM adds only 2.1 nm to the total dielectric thickness, it reduces the leakage current density between the plasma-treated aluminium bottom electrodes and thermally evaporated gold top electrodes by three orders of magnitude, from about 5×10^{-5} A cm^{-2} to $(5 \pm 1) \times 10^{-8}$ A cm^{-2} at an applied voltage of 2 V. Figure 1d shows that in comparison with the phosphonic acid-based dielectric, the plasma-grown oxide alone is a poor insulator with substantial leakage currents, small breakdown voltage, and significant device-to-device variation. (Low-voltage organic transistors employing a 5-nm-thick plasma-grown Al_2O_3 dielectric without SAM have been reported[13], but the gate leakage in these devices was $\sim 10^{-5}$ A cm^{-2} at 2 V.) Alkane-silanes also form SAMs on aluminium with large contact angles (100°), but the leakage currents are an order of magnitude larger compared with phosphonic acid SAMs (see Fig. 1d). For comparison, state-of-the-art low-power silicon transistors at the 90 nm technology node with an equivalent oxide thickness of 2.2 nm and a dielectric capacitance of 1.5 μF cm^{-2} have gate current densities around 10^{-3} A cm^{-2} (ref. 14).

The capacitance of our SAM-based dielectric was measured by impedance spectroscopy for frequencies between 1 Hz and 10 kHz for capacitors with contact areas between 0.28 and 1.13 cm^2. The measured capacitance is 0.7 ± 0.05 μF cm^{-2} (see Supplementary Fig. 2). This is approximately the value we calculate if we assume a permittivity of 2.5 for the SAM[15] and of 9 for the aluminium oxide[16], and a thickness of 2.1 nm for the SAM and of 3.6 nm for the aluminium oxide (as indicated by reflectivity measurements on Al/AlO_x/SAM samples, see Supplementary Information; $C_{total} = 1/[1/C_{SAM} + 1/C_{AlOx}]$).

X-ray photoelectron spectroscopy (XPS) was used to evaluate the surface composition. Spectra taken of an aluminium-coated silicon wafer before and after plasma treatment show Al (2*p*) and O (1*s*) signals that can be correlated with Al and AlO_x, but do not show a Si (2*p*) signal, suggesting that the aluminium covers the wafer completely. A spectrum taken after the preparation of a phosphonic acid-based SAM shows a P (2*p*) signal at a binding energy of 133.3 eV. This is the same binding energy observed in a powder sample of n-octadecylphosphonic acid, suggesting that the phosphonic acid group is not significantly altered by the adsorption on the surface. From the XPS data we have calculated the surface coverage of the SAMs and found that the packing density of n-octadecylphosphonic acid SAMs on plasma-treated aluminium is greater by a factor of 2.5 than the packing density of n-octadecyltrichlorosilane on the same surface (4.6 versus 1.9 molecules per nm^2). This may explain the difference in leakage currents and the fact that n-alkylphosphonic acid SAMs provide excellent insulation even in the absence of an aromatic endgroup[6].

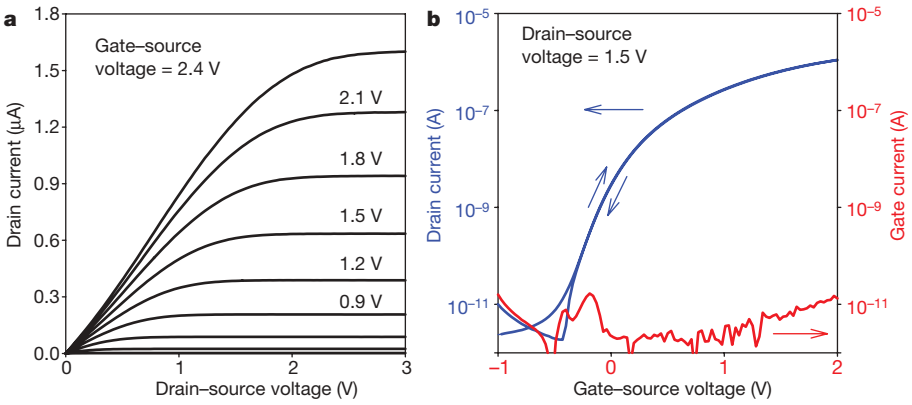


Figure 3 | n-channel $F_{16}CuPc$ TFT with SAM gate dielectric. The TFT has a channel length of 30 μm and a channel width of 1,000 μm. **a**, Drain current as a function of drain–source voltage. **b**, Drain current and gate current as a function of gate–source voltage. The maximum gate current is 15 pA and the carrier mobility is 0.02 $cm^2 V^{-1} s^{-1}$.

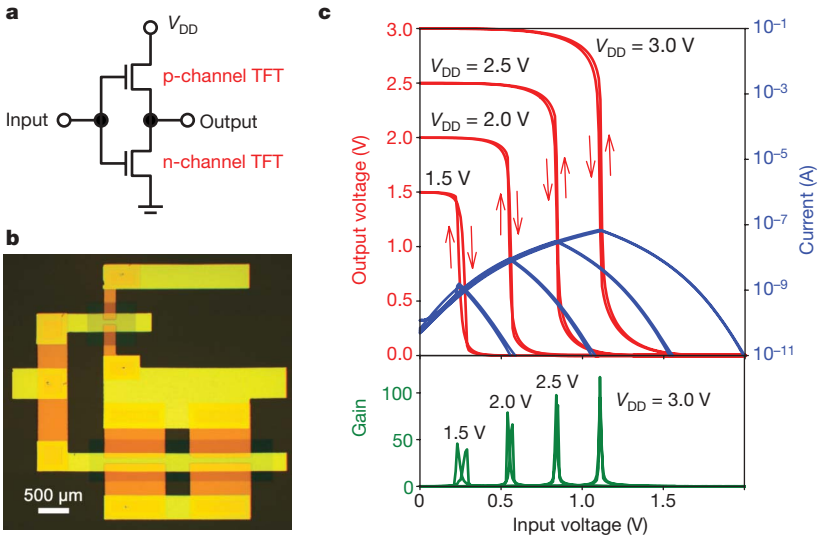


Figure 4 | Complementary inverter with SAM gate dielectric. **a**, Circuit schematic of the inverter. **b**, Photograph of the inverter. **c**, Output voltage, current, and small-signal gain as a function of input voltage for supply voltages between 1.5 and 3.0 V. The inverter shows rail-to-rail output switching, a maximum static current of 100 pA, and a small-signal gain as large as 100.

To prepare organic thin-film transistors (TFTs) and complementary circuits with SAM-based dielectric we have used a five-level shadow-mask process that is outlined in the Supplementary Information. The maximum process temperature is 90 °C. With the exception of the SAM growth, this is an all-dry manufacturing process that avoids the use of toxic or environmentally harmful solvents. We developed an integrated, one-step via process that exploits the specific surface selectivity of the self-assembling molecules during the adsorption process. All electrical measurements were carried out in ambient air at room temperature.

Figure 2 shows the electrical characteristics of a p-channel pentacene TFT with a channel length of 30 μm and a channel width of 100 μm. It has a carrier mobility of 0.6 $cm^2 V^{-1} s^{-1}$, a transconductance of 2.2 μS, a subthreshold swing of 100 mV per decade (see Supplementary Fig. 3), an on/off current ratio of 10^7, and a maximum gate current of 15 pA (see Supplementary Fig. 4 for statistics on 100 devices). These parameters are all similar to those reported previously for SAM gate dielectric TFTs on silicon substrates[6], but here this performance is demonstrated for TFTs with patterned metal gates on glass substrates. The gate current is smaller than the drain current by more than five orders of magnitude, confirming the high quality of the molecular gate dielectric. The mobility of the SAM-dielectric pentacene TFTs is similar to that of hydrogenated amorphous silicon TFTs widely used in commercial flat panel displays. We attribute this large mobility at least in part to the low surface energy[17] and low permittivity[18] of the molecular dielectric. Our organic TFTs can be operated with lower voltages than amorphous silicon TFTs and allow the realization of high-performance complementary devices and circuits.

The electrical characteristics of an n-channel $F_{16}CuPc$ TFT having a channel length of 30 μm, a channel width of 1,000 μm, a mobility of 0.02 $cm^2 V^{-1} s^{-1}$, a subthreshold swing of 160 mV per decade, an on/off ratio of 10^5, and a maximum gate current of 15 pA are shown in Fig. 3 (see also Supplementary Fig. 5). The mobility is similar to the best reported for $F_{16}CuPc$ TFTs on inorganic and on SAM-functionalized dielectrics[19–21].

Using p-channel pentacene and n-channel $F_{16}CuPc$ TFTs we have realized the first low-voltage organic complementary circuits with patterned gates on glass substrates. (Low-voltage organic complementary inverters have been demonstrated using a silicon gate and a 10- to 20-nm-thick polymer dielectric or a 100-nm-thick thermally grown SiO_2 dielectric; see refs 22, 23.) Our inverters with SAM-based gate dielectric show sharp switching with rail-to-rail output swings, large signal gain (~100), and negligible hysteresis for supply voltages as low as 1.5 V (see Fig. 4). Complementary two-input NAND gates also show the correct logic function (see Supplementary Fig. 6). Compared with circuits based on a single carrier type[2–4], complementary circuits consume far less static power, because all transistors of one carrier type are always in the non-conducting off-state, except during switching[24]. Indeed, the static currents (when $V_{in} = 0$ or $V_{in} = V_{DD}$) in our complementary gates are very small: always below 100 pA. Thus, the static power dissipation is less than 1 nW per logic gate, making these truly low-power organic circuits.

Complementary five-stage ring oscillators show stable oscillations for supply voltages as low as 1.5 V, the lowest operating voltage reported for an organic circuit. The circuits oscillate with rail-to-rail output voltage and signal delays as low as 1.4 ms per stage (see Supplementary Fig. 7). Improvements in circuit speed are expected by employing air-stable n-channel organic semiconductors with mobilities similar to pentacene[25] and TFTs with smaller critical dimensions (see Supplementary Fig. 8). All measurements were performed in air without protecting the circuits from ambient oxygen and humidity (see Supplementary Fig. 10 for results of operational stability tests and Supplementary Fig. 11 for shelf-life data).

Finally, we have taken advantage of the simplicity, robustness and low thermal budget of our process and prepared SAM-dielectric devices on flexible, transparent polyethylene naphthalate (PEN). Gate leakage is somewhat larger and carrier mobility is slightly smaller than on glass (see Supplementary Fig. 12), but this is expected when advancing from rigid to flexible substrates[26] and does not preclude the realization of low-power organic circuits on flexible substrates.

Received 21 June; accepted 14 December 2006.

1. Zhou, L. *et al.* All-organic active matrix flexible display. *Appl. Phys. Lett.* **88**, 083502 (2006).
2. Gelinck, G. H. *et al.* Flexible active-matrix displays and shift registers based on solution-processed organic transistors. *Nature Mater.* **3**, 106–110 (2004).
3. Someya, T. *et al.* Conformable, flexible, large-area networks of pressure and thermal sensors with organic transistor active matrixes. *Proc. Natl Acad. Sci. USA* **102**, 12321–12325 (2005).
4. Baude, P. *et al.* Pentacene-based radio-frequency identification circuitry. *Appl. Phys. Lett.* **82**, 3964–3966 (2003).
5. Collet, J., Tharaud, O., Chapoton, A. & Vuillaume, D. Low-voltage, 30 nm channel length, organic transistors with a self-assembled monolayer as gate insulating films. *Appl. Phys. Lett.* **76**, 1941–1943 (2000).
6. Halik, M. *et al.* Low-voltage organic transistors with an amorphous molecular gate dielectric. *Nature* **431**, 963–966 (2004).

7. Yoon, M. H., Facchetti, A. & Marks, T. J. σ-π molecular dielectric multilayers for low-voltage organic thin-film transistors. *Proc. Natl Acad. Sci. USA* **102,** 4678–4682 (2005).
8. Park, Y. D. *et al.* Low-voltage polymer thin-film transistors with a self-assembled monolayer as the gate dielectric. *Appl. Phys. Lett.* **87,** 243509 (2005).
9. Ulman, A. *An Introduction to Ultrathin Films: From Langmuir-Blodgett to Self-Assembly* (Academic Press, Boston, 1991).
10. Folkers, J. P., Gorman, C. B., Laibinis, P. E., Buchholz, S. & Whitesides, G. M. Self-assembled monolayers of long-chain hydroxamic acids on the native oxides of metals. *Langmuir* **11,** 813–824 (1995).
11. Bram, C., Jung, C. & Stratmann, M. Self assembled molecular monolayers on oxidized inhomogeneous aluminum surfaces. *Fresenius J. Anal. Chem.* **358,** 108–111 (1997).
12. Goetting, L. B., Deng, T. & Whitesides, G. M. Microcontact printing of alkanephosphonic acids on aluminum: Pattern transfer by wet chemical etching. *Langmuir* **15,** 1182–1191 (1999).
13. Kim, K. D. & Song, C. K. Low voltage pentacene thin film transistors employing a self-grown metal-oxide as a gate dielectric. *Appl. Phys. Lett.* **88,** 233508 (2006).
14. Wu, C. C. *et al.* A 90-nm CMOS device technology with high-speed, general-purpose, and low-leakage transistors for system on chip applications. *2002 International Electron Devices Meeting Technical Digest* 65–68 (The Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, 2002).
15. Fontaine, P. *et al.* Octadecyltrichlorosilane monolayers as ultrathin gate insulating films in metal-insulator-semiconductor devices. *Appl. Phys. Lett.* **62,** 2256–2258 (1993).
16. Robertson, J. High dielectric constant oxides. *Eur. Phys. J. Appl. Phys.* **28,** 265–291 (2004).
17. Yang, S. Y., Shin, K. & Park, C. E. The effect of gate dielectric surface energy on pentacene morphology and organic field-effect transistor characteristics. *Adv. Funct. Mater.* **15,** 1806–1814 (2005).
18. Hulea, I. N. *et al.* Tunable Fröhlich polarons in organic single-crystal transistors. *Nature Mater.* **5,** 982–986 (2006).
19. Bao, Z., Lovinger, A. & Brown, J. New air-stable n-channel organic thin film transistors. *J. Am. Chem. Soc.* **120,** 207–208 (1998).
20. de Oteyza, D. G. *et al.* Controlled enhancement of the electron field-effect mobility of F_{16}CuPc thin-film transistors by use of functionalized SiO_2 substrates. *Appl. Phys. Lett.* **87,** 183504 (2005).
21. Ling, M. M. & Bao, Z. Copper hexafluorophthalocyanine field-effect transistors with enhanced mobility by soft contact lamination. *Org. Electron.* **7,** 568–575 (2006).
22. Yoon, M. H., Yan, H., Facchetti, A. & Marks, T. J. Low-voltage organic field-effect transistors and inverters enabled by ultrathin cross-linked polymers as gate dielectrics. *J. Am. Chem. Soc.* **127,** 10388–10395 (2005).
23. De Vusser, S., Steudel, S., Myny, K., Genoe, J. & Heremans, P. Low voltage complementary organic inverters. *Appl. Phys. Lett.* **88,** 162116 (2006).
24. Crone, B. K. *et al.* Design and fabrication of organic complementary circuits. *J. Appl. Phys.* **89,** 5125–5132 (2001).
25. Jones, B. A. *et al.* High-mobility air-stable n-type semiconductors with processing versatility: Dicyanoperylene-3,4:9,10-bis(dicarboximides). *Angew. Chem. Int. Edn Engl.* **43,** 6363–6366 (2004).
26. Klauk, H. *et al.* Pentacene organic transistors and ring oscillators on glass and on flexible polymeric substrates. *Appl. Phys. Lett.* **82,** 4175–4177 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Partial financial support (J.P. and M.H.) was provided by the Deutsche Forschungsgemeinschaft.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.K. (h.klauk@fkf.mpg.de).

LETTERS

Charge- and size-based separation of macromolecules using ultrathin silicon membranes

Christopher C. Striemer[1], Thomas R. Gaborski[2], James L. McGrath[2] & Philippe M. Fauchet[1]

Commercial ultrafiltration and dialysis membranes have broad pore size distributions and are over 1,000 times thicker than the molecules they are designed to separate, leading to poor size cut-off properties, filtrate loss within the membranes, and low transport rates[1,2]. Nanofabricated membranes have great potential in molecular separation applications by offering more precise structural control[3,4], yet transport is also limited by micrometre-scale thicknesses[5]. This limitation can be addressed by a new class of ultrathin nanostructured membranes where the membrane is roughly as thick (~10 nm) as the molecules being separated, but membrane fragility and complex fabrication have prevented the use of ultrathin membranes for molecular separations[1]. Here we report the development of an ultrathin porous nanocrystalline silicon (pnc-Si) membrane using straightforward silicon fabrication techniques that provide control over average pore sizes from approximately 5 nm to 25 nm. Our pnc-Si membranes can retain proteins while permitting the transport of small molecules at rates an order of magnitude faster than existing materials, separate differently sized proteins under physiological conditions, and separate similarly sized molecules carrying different charges. Despite being only 15 nm thick, pnc-Si membranes that are free-standing over 40,000 μm^2 can support a full atmosphere of differential pressure without plastic deformation or fracture. By providing efficient, low-loss macromolecule separations, pnc-Si membranes are expected to enable a variety of new devices, including membrane-based chromatography systems and both analytical and preparative microfluidic systems that require highly efficient separations.

Given the potential of nanofabricated membranes to advance macromolecular separation processes and the limitations of existing materials, we have developed a robust and inexpensive ultrathin porous membrane. Our fabrication process is shown schematically in Fig. 1a, and is described in detail in the Methods section. Briefly, our approach uses precision silicon deposition and etching techniques to create the ultrathin membrane. However, instead of directly patterning pores, we have discovered that voids are formed spontaneously as nanocrystals nucleate and grow in a 15-nm-thick amorphous silicon (a-Si) film during a rapid thermal annealing step. The voids span the molecularly thin membrane to create pores. The resulting membranes cover openings several hundred micrometres across in a rigid crystalline silicon frame, and can therefore be easily handled and used.

Figure 1b shows a plan-view transmission electron microscopy (TEM) image of a 15-nm-thick pnc-Si membrane. In this image, the pores appear as white circles while the solid nanocrystalline silicon is grey and black owing to electron diffraction. Nanocrystals with aligned atomic planes that satisfy the Bragg condition for the electron beam cause strong diffraction, and make the nanocrystals appear black in the bright field image. Nanocrystals not satisfying the Bragg condition appear grey. All structures visible in this image were formed during a single 30 s anneal at 770 °C, yielding well defined pores with diameters of 9–35 nm.

In addition to TEM, several other characterization techniques have been used to confirm the properties of our pnc-Si membranes. Figure 2a shows refractive index dispersion data obtained using spectroscopic ellipsometry for a 15-nm-thick silicon film after deposition

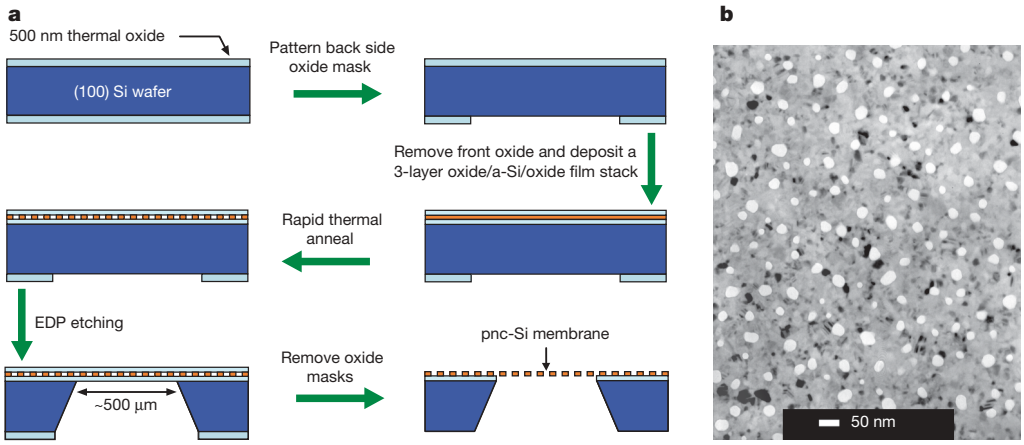


Figure 1 | Fabrication and pore morphology of pnc-Si membrane. **a**, Standard microfabrication tools are used to deposit, pattern and suspend pnc-Si films. **b**, A plan-view TEM image of the porous nanostructure of a 15-nm-thick membrane. In this bright field image, pores appear as bright spots while nanocrystalline silicon is in grey or black contrast.

[1]Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627, USA. [2]Department of Biomedical Engineering, University of Rochester, Rochester, New York 14620, USA.

(a-Si) and after crystallization (pnc-Si). Our sputtered a-Si has high optical density, comparable to microelectronic quality a-Si deposited with chemical vapour deposition, and exhibits a clear shift in optical properties after crystallization, with characteristic resonance peaks similar to crystalline silicon[6]. These data are indicative of high purity silicon films with smooth interfaces. It should also be noted that TEM images of the as-deposited a-Si show no distinguishable voids or crystalline features. To confirm the accuracy of our spectroscopic ellipsometry data, several membranes were transferred onto polished quartz and atomic force microscopy (AFM) was used to measure the step height of the membrane edge. Figure 2b shows an AFM image

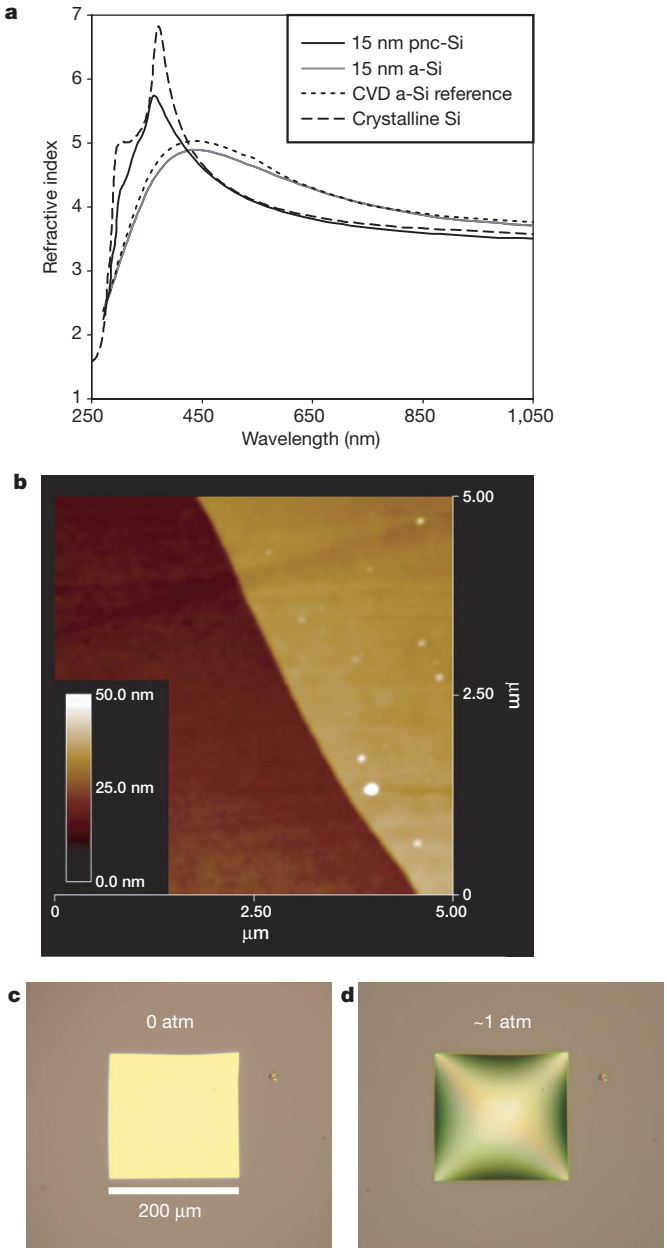


Figure 2 | Physical properties of pnc-Si membranes. a, Refractive index dispersion curves of our high optical density silicon film before and after crystallization (729 °C, 30 s anneal), determined by spectroscopic ellipsometry. Dispersion curves for crystalline silicon and a-Si grown by chemical vapour deposition (CVD) are also plotted for reference. **b**, An AFM scan over the edge of a membrane transferred to a polished quartz window confirms 15 nm thickness and minimal roughness. **c**, **d**, Optical micrographs of a 15-nm-thick pnc-Si membrane at equilibrium (**c**) and with ~1 atm (15 p.s.i.) of back pressure (**d**), demonstrating the remarkable strength of this ultrathin material.

that confirms the 15 nm thickness of a sample membrane and its highly smooth surface morphology.

Another important characteristic of pnc-Si membranes is their remarkable mechanical stability. We mechanically tested membranes using a customized holder to apply pressure to one side of the membrane while an optical microscope was used to monitor deformation. Figure 2c, d shows optical micrographs of a 200 μm × 200 μm × 15 nm membrane as more than 1 atm of differential pressure was applied across it for approximately 5 min. With no differential pressure, the membrane is extremely flat (Fig. 2c), and at maximum pressure (Fig. 2d) the membrane elastically deforms but maintains its structural integrity throughout the duration of test. Unlike thin polymer membranes[7], pnc-Si membranes exhibit no plastic deformation and immediately return to their flat state when the pressure is removed. Pressurization tests were cycled three times with no observable membrane degradation. The remarkable strength and durability exhibited by these membranes is likely to be due to their smooth surfaces[1] and random nanocrystal orientation that inhibit the formation and propagation of cracks.

We have also determined that pore size distributions in pnc-Si membranes can be controlled through adjustment of the rapid thermal annealing temperature during crystallization. Nanocrystal nucleation and growth are Arrhenius-like processes[8] that exhibit strong temperature dependence above a threshold crystallization temperature of approximately 700 °C in a-Si (ref. 9). Existing crystallization models[10] fail to predict void formation, and must be extended to account for how volume contraction and material strain lead to pore formation in ultrathin membranes. To demonstrate pore size tunability, three wafers with 15-nm-thick pnc-Si membranes were processed identically, except for the annealing temperature. TEM images of these membranes (Fig. 3, right) revealed that pore size and density increase monotonically with temperature, as samples annealed at 715 °C, 729 °C and 753 °C have average pore sizes of 7.3 nm, 13.9 nm and 21.3 nm, respectively. A sample annealed at 700 °C exhibited no crystallinity or voids, illustrating a strong morphological dependence on temperature near the onset of crystallization. The tunability of pore size in this range makes pnc-Si membranes particularly well suited for size-selective separation of large biomolecules, such as proteins and DNA. Because pore area is central to the discussion of molecular transport through these membranes, histograms that identify the total pore area available at each pore size are presented in Fig. 3, left. Pore size data were extracted directly from the micrographs using image processing software.

To demonstrate molecular separations with pnc-Si, we chose two common blood proteins of different molecular weight (MW) and hydrodynamic diameter (D); bovine serum albumin, BSA (MW = 67,000 (67K), $D = 6.8$ nm) and immunoglobulin-γ, IgG (MW = 150K, $D = 14$ nm), fluorescently labelled with Alexa 488 and Alexa 546 (Molecular Probes), respectively. Free Alexa 546 dye was used as an additional low molecular weight (MW = 1K, $D \approx 1$ nm) species. We monitored the passage of these species through the pnc-Si membranes shown in Fig. 3 using real time fluorescence microscopy (Fig. 4a). In this set-up, a membrane and its supporting silicon wafer frame was placed above a glass slide with 50 μm silica bead spacers, forming a thin diffusion chamber beneath the membrane. The chamber was first filled with approximately 50 μl of clean buffer solution and then 3 μl of a fluorescent mixture was added to the well above the membrane. An image of the membrane edge was taken every 30 s in each of the fluorescence channels. The passage of each species through the membrane was observed as the spreading of fluorescence signal from the membrane edge, as illustrated in the two false-colour images in Fig. 4a.

Figure 4b shows quantitative results for the separation of free Alexa 546 dye and BSA using membrane A (Fig. 3a). These data were generated by plotting the fluorescence intensity at a point 50 μm away from the membrane edge for a time series of images. From these results, it is clear that dye passes freely through the membrane while

BSA is almost completely blocked. Figure 4c shows a similar experiment, where the permeabilities of IgG and BSA through membrane B at 1 μM concentration are compared. In this case, BSA diffuses through the membrane >4 times more rapidly than IgG. Because the diffusion coefficients for these molecules are within 25% of each other[11], the rate difference that we measure clearly indicates that pnc-Si membranes hinder IgG diffusion relative to BSA diffusion. By more thoroughly optimizing pore sizes, we expect to engineer pnc-Si membranes that can completely exclude IgG but permit BSA passage. The plots of Fig. 4b and c can be quantitatively compared for BSA, demonstrating that the increased cut-off size of membrane B allows a 15 times enhancement of BSA diffusion relative to membrane A. Time-lapse movies of these two separation experiments are included in Supplementary Information.

An interesting finding from our separation experiments was that BSA and IgG are retained behind membranes with maximal pore sizes more than twice as large as their reported hydrodynamic diameters. As a potential explanation, we investigated the possibility that electrostatic interactions and protein adsorption might create an effective pore size smaller than that measured by TEM. Indeed, the passage of negatively charged Alexa 488 (2−) dye in the presence and absence of high salt concentrations suggests a role for electrostatic interactions in determining effective pore sizes. As shown in Fig. 4d, the transport of the Alexa dye drops by a factor of 10 when experiments are conducted in deionized water. We interpret this as a consequence of electrostatic repulsion between the dye and a negatively charged native oxide layer on the surface of the pnc-Si membranes.

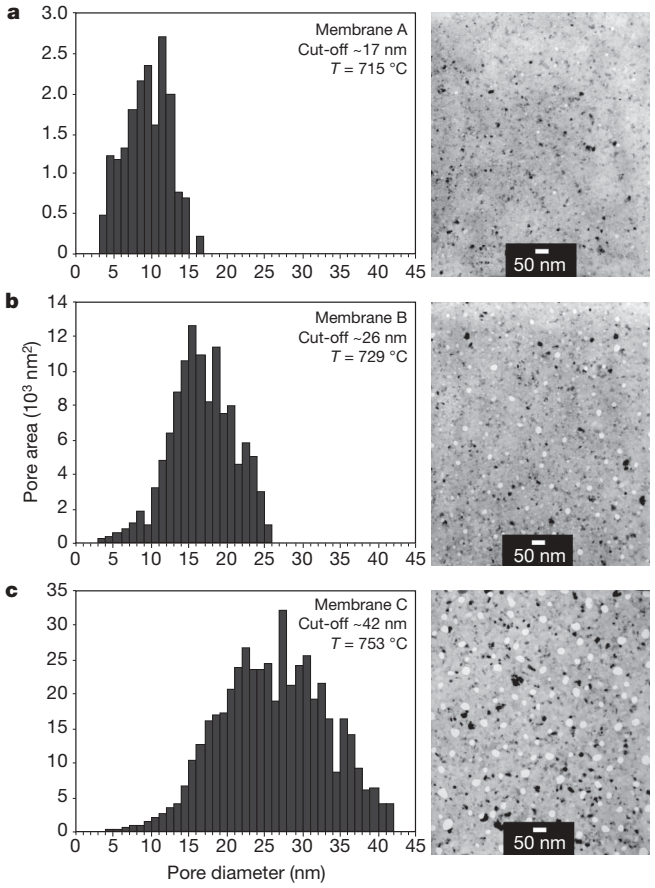


Figure 3 | Tunability of pnc-Si membrane pore size for molecular separation experiments. By varying the temperature at which the silicon film is crystallized, pore diameter can be controlled. The maximum (cut-off) pore size and porosity increase with annealing temperature from **a**, 715 °C to **b**, 729 °C and **c**, 753 °C, as illustrated in the histograms that plot total pore area available for molecular transport at each pore diameter. The TEM image used to generate each histogram is included on the right side.

Consistent with this interpretation, high salt concentrations increase throughput by screening surface and solute charges. To provide further evidence of charge effects, we modified membranes to carry abundant negative and positive surface charges. We found that in low ionic strength solutions, positively charged membranes blocked only positively charged dyes and negatively charged membranes blocked only negatively charged dyes (Supplementary Fig. S1). Although the high ionic strength of phosphate buffered saline should largely screen the charge of the native membrane oxide for our protein separations, proteins are much larger molecules and can be heavily charged (BSA net molecular charge is 13− at pH 7)[12], so stronger electrostatic interactions that reduce the effective pore size are expected. In addition to offering a partial explanation for reduced effective pore sizes in charge separations, these studies clearly demonstrate that pnc-Si membranes can be functionalized to separate similarly sized molecules on the basis of their charge.

Protein adsorption to the pore walls will also reduce effective pore size. By directly staining for protein on pnc-Si membranes used for separations, it is evident that BSA adsorption shrinks, but does not occlude, the largest membrane pores by as much as 7 nm (Supplementary Fig. S2). In addition to charge and adsorption effects, we expect other factors to reduce effective pore sizes, such as the uncertain relationship between a protein's physical size and hydrodynamic dimensions, and the behaviour of water (hydrogen bonding) in nanoscale pores.

Given the hours-long passage-times of molecules through thick membranes[3], it is significant that filtrate molecules appear downstream of pnc-Si filters within minutes. To better quantify the transport through pnc-Si membranes, we followed the fluorescence microscopy experiments with bench-top experiments in which we could remove and assay the Alexa 546 dye that diffused across membrane A from a 100 μM starting concentration using a similar unstirred geometry. We compared dye diffusion through pnc-Si membranes to diffusion through standard regenerated cellulose dialysis membranes (Spectra/Por 7 dialysis membrane, molecular-weight cut-off = 50K). The results shown in Fig. 4e reveal that dye diffuses over 9 times faster through pnc-Si membranes than dialysis membranes with comparable size exclusion properties (a 50K cut-off dialysis membrane was chosen, based on the excellent retention of BSA (67K) by membrane A in Fig. 4b). The pnc-Si membranes exhibit an initial transport rate of 156 nmol $cm^{-2} h^{-1}$ (Fig. 4e) that rapidly slows as the 3 μl source volume depletes, lowering the concentration gradient across the barrier. Remarkably, when this experiment was repeated with membrane C for 1 h, an increase of <10% in dye transport was measured relative to membrane A, despite porosities differing by ~29 times (0.2% versus 5.7%). We interpret this as evidence that dye or small molecule transport is essentially unhindered by our membranes, as porosities far lower than that of membrane A should theoretically allow greater than half-maximal diffusion through an infinitely thin porous barrier[13]. Therefore, the observed 9 times increase in diffusion rate over conventional dialysis membranes indicates that diffusion through the commercial membrane is the rate-limiting transport process, whereas diffusion through the bulk solution is rate-limiting for the pnc-Si membrane experiment. Substantial enhancement of transport rate is expected in systems that implement active mixing, or forced flow (pressure- or voltage-driven).

Our work with pnc-Si membranes represents the first use of ultrathin nanomembranes for size-based molecular separations, and encourages their use in several near-term applications. First, the separation of BSA and IgG suggests that pnc-Si can be used for membrane-based protein fractionation. BSA and IgG are too close in size (2.2 times MW difference) to be efficiently separated using conventional membrane processes[14] (>10 times MW difference is recommended by the manufacturers), and much of the filtrate species is lost to the high surface area and tortuous porosity of these standard membranes. By minimizing filter material, pnc-Si

membranes should allow for recovery of both the retentate and filtrate fractions to enable membrane-based chromatography. Second, because they are molecularly thin and have a minimal filter surface area, pnc-Si membranes are expected to be highly efficient for separation processes. Indeed, our diffusion measurements recorded a transport rate of 156 nmol cm^{-2} h^{-1} for Alexa 546 dye. This rate is more than one order of magnitude faster than those for thick nanofabricated membranes[3], and >9 times faster than our own measurements through dialysis membranes. We further demonstrated that pnc-Si membranes with fixed charges can be used to separate similarly sized molecules with different charges, adding another dimension of control for highly efficient molecular separations[15]. Finally, the silicon-based platform opens several avenues for future developments, including scalable production of membranes, straightforward integration into microfluidic devices, and surface functionalization using well-established chemistries[16,17] to modify surface charge, reduce protein adsorption and protect the silicon from chemical attack in harsh environments. Importantly, the demonstrated

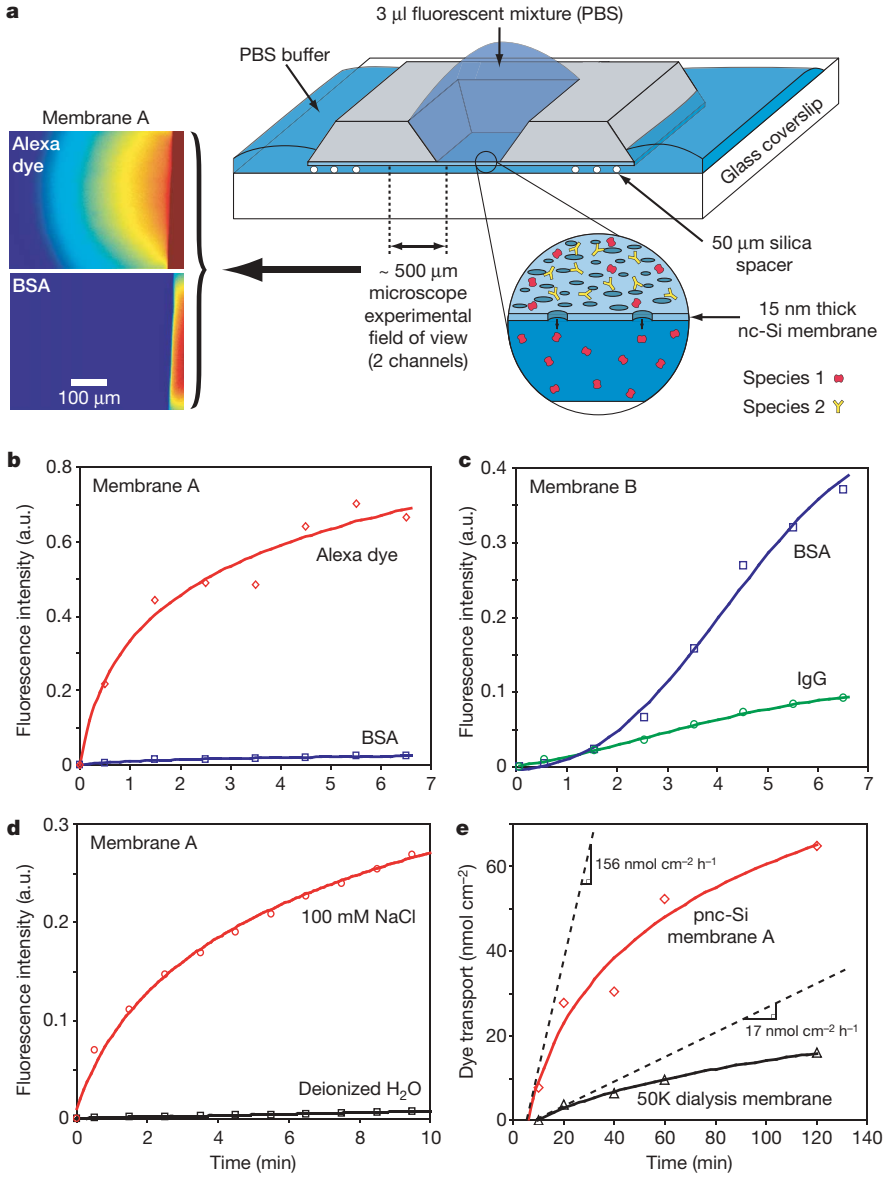


Figure 4 | Molecular separation and transport rates through pnc-Si membranes. **a**, The passage of two fluorescent species (labelled proteins or free dye) from a 3 µl source mixture through pnc-Si membranes was monitored simultaneously on two channels of a fluorescence microscope. The membrane edge was imaged from below and the lateral spread of fluorescent material was monitored to determine permeation through the membrane. Experimental false-colour images of the two fluorescence channels taken 6.5 min after the application of a mixture of labelled BSA (MW = 67K) and free Alexa 546 dye (MW = 1K) clearly show species separation. The membrane edge is the bright feature on the right side of each image, corresponding to the fluorescence generated in the source solution behind the membrane. Spread of the fluorescence signal beyond the membrane edge quantitatively indicates transport through the membrane. **b**, As shown in **a**, highly efficient separation of BSA and free dye was observed through membrane A (Fig. 3a) over 6.5 min. A plot of the fluorescence intensity 50 µm from the membrane edge was generated from a time series of images. Intensities were normalized to the centre value of the membrane in the first frame for each channel. **c**, A greater than fourfold separation of proteins BSA and IgG (MW = 150K) was observed through membrane B (Fig. 3b) using the same method. **d**, The transport of dye through membrane A was determined for solutions with high and low ionic strength using the set-up from **a**. **e**, The diffusion rate of dye through membrane A was benchmarked relative to a commercial dialysis membrane with a 50K cut-off (50 times larger than the dye MW). Because the 3 µl of source dye quickly depletes, the transport rate is calculated as the initial slope of the transport curve. Dye concentration was measured by 558 nm absorption at each time point.

mechanical strength of these ultrathin membranes should allow the construction of large-scale dialysis systems, and facilitate their use in pressurized filtration devices at the macro- and microscale.

METHODS

Fabrication. The pnc-Si membranes are fabricated by the procedure outlined in Fig. 1. We first grow a 500-nm-thick layer of SiO_2 on both sides of a silicon wafer. On the back side of the wafer, the SiO_2 is patterned using standard photolithography techniques to form an etch mask for the membrane formation process. The front oxide layer is then removed, and a high quality three layer film stack (20 nm SiO_2 / 15 nm a-Si / 20 nm SiO_2) is deposited on the front surface using RF magnetron sputtering. The a-Si layer is sputtered at a chamber pressure of 1.5 mtorr in Ar with a target power density of 0.4 W cm^{-2}, yielding a deposition rate of 3.4 nm min^{-1}. The SiO_2 layers are reactively sputtered from a silicon target at a chamber pressure of 1.5 mtorr, with a (3:4) Ar:O_2 gas flow ratio, and a target power density of 1.8 W cm^{-2}, yielding a deposition rate of 10.7 nm min^{-1}. Our deposition recipe is well characterized, and we are able to deposit films with $\pm 1\%$ thickness accuracy and surface roughness <0.5 nm. We have previously demonstrated the crystallization of very thin amorphous silicon films, forming high quality nanocrystals with well-defined size[18].

To form the pnc-Si membranes, the substrate is briefly exposed to high temperature (715–770 °C for 30 s) in a rapid thermal processing chamber, crystallizing the a-Si into a nanocrystalline film. The patterned wafer back side is then exposed to a highly selective silicon etchant, EDP (ethylenediamine pyrocatechol)[19], which removes the silicon wafer along (111) crystal planes until it reaches the first SiO_2 layer of the front side film stack. This etch is also used to outline approximately 80 samples (3.5 mm × 9 mm) that can be easily removed from the wafer after the fabrication process is complete, and used individually for molecular separation experiments. Finally, exposing the three layer membrane to buffered oxide etchant removes the protective oxide layers, leaving only the freely suspended ultrathin pnc-Si membrane. We have used this process to fabricate square membranes as thin as 5 nm and as large as 2 mm × 2 mm, but in this Letter we focus on more structurally robust 15-nm-thick membranes, measuring several hundred micrometres per side.

Membrane processing. Silicon surfaces exposed to air tend to grow a native oxide that reaches a thickness of ~1 nm within a few hours and stabilizes at ~1.5 nm in approximately 1 week (characterized by us using spectroscopic ellipsometry). This process can be accelerated by exposure to an oxygen plasma at 450 °C and the oxide growth also self-limits at ~1.5 nm thickness under these conditions. The membranes used in Supplementary Fig. S1 were exposed to a 500 W RF oxygen plasma at 450 °C and 12 mtorr for 60 s to form a stable surface oxide. Silanization was then performed on several membranes after oxidation, using a recipe that we have used previously[17]. Briefly, the membranes were first cleaned in a 1:1 solution of HCl:methanol for 30 min. The samples were then soaked for 15 min in a solution of 1.5 ml 5% (aminopropyl)triethoxysilane (APTES) in deionized water added to 28.5 ml acetone. Membranes were then dried and baked for 15 min at 100 °C. This process forms a monolayer of APTES with a thickness of approximately 0.8 nm (ref. 20).

Protein separation. For the protein separation experiments, BSA and IgG were labelled with Alexa Fluor 488 and 546 dyes (Molecular Probes). The dyes react with primary amines, forming stable covalent bonds. Each species was twice purified with spin columns provided in the labelling kit. Protein concentration and degree of labelling was calculated by measuring absorbances with a spectrophotometer using extinction coefficients provided by the dye manufacturer. This analysis showed that BSA was labelled with eight moles of dye per mole of protein, while IgG was labelled with three moles of dye per mole of protein. In our microscope, this yielded similar fluorescence intensity for each species at the same concentration. In the protein separation experiments, proteins were used at 1 μM, while free dye was used at 100 μM to mimic the separation of proteins from higher concentration solute species, as might occur in buffer exchange or desalting applications. In the electrochemical double layer experiments, Alexa and Rhodamine 6G dyes were used at 50 μM. All plots and figures were normalized (as described in Fig. 4 legend) for comparative purposes.

Received 23 March; accepted 20 November 2006.

1. Tong, H. D. *et al.* Silicon nitride nanosieve membrane. *Nano Lett.* **4,** 283–287 (2004).
2. Kuiper, S., van Rijn, C. J. M., Nijdam, W. & Elwenspoek, M. C. Development and applications of very high flux microfiltration membranes. *J. Membr. Sci.* **150,** 1–8 (1998).
3. Yamaguchi, A. *et al.* Self-assembly of a silica-surfactant nanocomposite in a porous alumina membrane. *Nature Mater.* **3,** 337–341 (2004).
4. Lee, S. B. & Martin, C. R. Electromodulated molecular transport in gold-nanotubule membranes. *J. Am. Chem. Soc.* **124,** 11850–11851 (2002).
5. Martin, F. *et al.* Tailoring width of microfabricated nanochannels to solute size can be used to control diffusion kinetics. *J. Control. Release* **102,** 123–133 (2005).
6. Palik, E. D. *Handbook of Optical Constants of Solids* 547–569 (Academic, Orlando, 1985).
7. Jiang, C. *et al.* Freely suspended nanocomposite membranes as highly sensitive sensors. *Nature Mater.* **3,** 721–728 (2004).
8. Spinella, C., Lombardo, S. & Priolo, F. Crystal grain nucleation in amorphous silicon. *J. Appl. Phys.* **84,** 5383–5414 (1998).
9. Zacharias, M. *et al.* Thermal crystallization of amorphous Si/SiO_2 superlattices. *Appl. Phys. Lett.* **74,** 2614–2616 (1999).
10. Zacharias, M., Heitmann, J., Schmidt, M. & Streitenberger, P. Confinement effects in crystallization and Er doping of Si nanostructures. *Physica E* **11,** 245–251 (2001).
11. Karlsson, D., Zacchi, G. & Axelsson, A. Electronic speckle pattern interferometry: a tool for determining diffusion and partition coefficients for proteins in gels. *Biotechnol. Prog.* **18,** 423–430 (2002).
12. Pujar, N. S. & Zydney, A. L. Electrostatic effects on protein partitioning in size-exclusion chromatography and membrane ultrafiltration. *J. Chromatogr. A* **796,** 229–238 (1998).
13. Berg, H. C. *Random Walks in Biology* 34–37 (Princeton Univ. Press, Princeton, New Jersey, 1993).
14. Li, Q. Y., Cui, Z. F. & Pepper, D. S. Fractionation of HSA and IgG by gas sparged ultrafiltration. *J. Membr. Sci* **136,** 181–190 (1997).
15. Ku, J.-R. & Stroeve, P. Protein diffusion in charged nanotubes: ''On-off'' behavior of molecule transport. *Langmuir* **20,** 2030–2032 (2004).
16. Letant, S. E., Hart, B. R., Van Buuren, A. W. & Terminello, L. J. Functionalized silicon membranes for selective bio-organism capture. *Nature Mater.* **2,** 391–395 (2003).
17. Mace, C. R., Striemer, C. C. & Miller, B. L. Theoretical and experimental analysis of arrayed imaging reflectometry as a sensitive proteomics technique. *Anal. Chem.* **78,** 5578–5583 (2006).
18. Grom, G. F. *et al.* Ordering and self-organization in nanocrystalline silicon. *Nature* **407,** 358–361 (2000).
19. Reisman, A. *et al.* The controlled etching of silicon in catalyzed ethylenediamine-pyrocatechol-water solutions. *J. Electrochem. Soc.* **126,** 1406–1415 (1979).
20. Ouyang, H., Striemer, C. C. & Fauchet, P. M. Quantitative analysis of the sensitivity of porous silicon optical biosensors. *Appl. Phys. Lett.* **88,** 163108–163110 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Krishnan for assistance with ellipsometry and AFM, and for useful discussions; J. Snyder for assistance in preparing protein-treated membrane samples for TEM; B. McIntyre for assistance with microscopy; P. Osborne for help with equipment fabrication; and the Laboratory for Laser Energetics at the University of Rochester for providing access to their spectroscopic ellipsometer. Silicon microprocessing was conducted at the Hopeman Microfabrication Facility at the University of Rochester and the Semiconductor and Microsystems Fabrication Laboratory (SMFL) at the Rochester Institute of Technology. This work was partially supported by a Johnson & Johnson Award to the University of Rochester Medical School's Discovery Concept Fund.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.C.S. (striemer@ece.rochester.edu).

LETTERS

The Earth's 'hum' is driven by ocean waves over the continental shelves

Spahr C. Webb[1]

Observations show that the seismic normal modes of the Earth at frequencies near 10 mHz are excited at a nearly constant level in the absence of large earthquakes[1]. This background level of excitation has been called the 'hum' of the Earth[2], and is equivalent to the maximum excitation from a magnitude 5.75 earthquake[3]. Its origin is debated, with most studies attributing the forcing to atmospheric turbulence, analogous to the forcing of solar oscillations by solar turbulence[2,4–7]. Some reports also predicted that turbulence might excite the planetary modes of Mars to detectable levels[4]. Recent observations on Earth, however, suggest that the predominant excitation source lies under the oceans[8–10]. Here I show that turbulence is a very weak source, and instead it is interacting ocean waves over the shallow continental shelves that drive the hum of the Earth. Ocean waves couple into seismic waves through the quadratic nonlinearity of the surface boundary condition, which couples pairs of slowly propagating ocean waves of similar frequency to a high phase velocity component at approximately double the frequency. This is the process by which ocean waves generate the well known 'microseism peak' that dominates the seismic spectrum near 140 mHz (refs 11, 12), but at hum frequencies, the mechanism differs significantly in frequency and depth dependence. A calculation of the coupling between ocean waves and seismic modes reproduces the seismic spectrum observed. Measurements of the temporal correlation between ocean wave data and seismic data[9,10] have confirmed that ocean waves, rather than atmospheric turbulence, are driving the modes of the Earth.

Observations of the normal mode spectrum of the Earth made in the absence of large earthquakes show a roughly constant level, except for a small biannual cycle with energy peaking in January and July, which is consistent with the most energetic storm seasons (and hence largest ocean waves) in the Northern and Southern Hemispheres, respectively[3,6]. The modes appear as a series of lines between 1 and 10 mHz in spectra from quiet seismometer sites during days without large earthquakes (Fig. 1a). Above 10 mHz, distinct lines are not resolved and the Earth's hum is better described as propagating Rayleigh waves[5]. The many small earthquakes that occur each day provide insufficient seismic moment to explain quiet day spectra[1,6].

It has long been known that ocean waves drive the large 'microseism' peak in the worldwide seismic noise spectrum near 140 mHz (refs 11, 12; Fig. 1a). Components of the wave field interact to generate components that force seismic waves because ocean waves are weakly nonlinear[11]. Schematically, two ocean waves of frequencies ω_1, ω_2 and horizontal wavenumbers $\mathbf{k}_1$, $\mathbf{k}_2$, interact to produce a signal at frequency $\omega_3 = \omega_1 + \omega_2$ and wavenumber $\mathbf{k}_3 = \mathbf{k}_1 + \mathbf{k}_2$. If two waves with $\omega_1 \approx \omega_2$ are travelling in opposing directions, so that $\mathbf{k}_1 \approx -\mathbf{k}_2$, then $|\mathbf{k}_3| \ll |\mathbf{k}_1|$ and $\mathbf{k}_3$ corresponds to forcing at a phase speed $c_3 = \omega/|\mathbf{k}_3|$, much larger than the speed of the original waves. Thus energy in 14 s period (70 mHz) ocean waves ($c < 20\,\mathrm{m\,s^{-1}}$) is transferred to seismic waves ($c > 1.5\,\mathrm{km\,s^{-1}}$) near 7 s period (140 mHz) that travel worldwide to generate the microseism peak (Fig. 1a). Similarly, low frequency ocean waves interact over the shelves to force Earth normal modes and Rayleigh waves below 40 mHz. A 5 mHz fundamental Earth seismic mode is of 900 km wavelength[13], is described by spherical harmonics of angular order $l \approx 43$ ($c > 4.5\,\mathrm{km\,s^{-1}}$), and is forced by ocean waves at 2.5 mHz.

The finite extent of regions of strong winds limits the longest period ocean waves driven directly by the wind to periods shorter than 25 s. The much smaller amplitude ocean waves observed at longer period (relevant to seismic mode excitation) are called 'infragravity waves'. These waves are most energetic near the shore, and are driven by a nonlinear mechanism related to the microseism mechanism acting on short period ocean waves near coastlines. The infragravity wave spectrum on the shelves[14,15] varies greatly both spatially and in time, but remains flat in frequency from 1 mHz to 40 mHz (above which wind waves dominate, Fig. 1b). Infragravity wave spectra are typically 40 dB more energetic over the shelf than in the deep ocean because little wave energy reaches deep water (the sloping shelf acts as a waveguide).

Hasselmann[11] derived a small wavenumber approximation for the wavenumber and frequency spectrum $F_p(\mathbf{k},\omega)$ of the near surface pressure field in water of infinite depth driven by the nonlinear

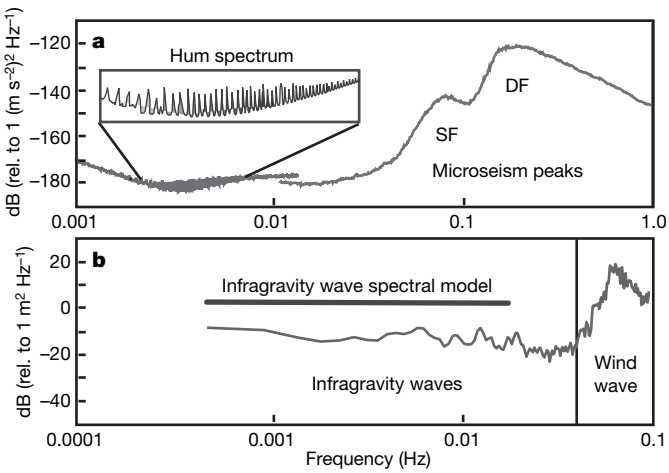


Figure 1 | Seismic waves are driven by ocean waves at half their frequency. **a**, Vertical acceleration spectrum from a quiet site (BFO, Black Forest Observatory), redrawn from data supplied by R. Widmer-Schnidrig (available at http://www-gpi.physik.uni-karlsruhe.de/pub/widmer/BFO/Noise/BFO_STS-1_BHZ_VHZ.pdf). Normal mode spectral peaks (Earth's hum) lie between 1 and 10 mHz, and are shown magnified in the inset. The DF microseism peak is driven by ocean waves near 70 mHz, the hum by lower frequency ocean waves. The 'SF' peak is probably driven by waves interacting with bathymetry[11]. **b**, Ocean wave height spectrum from the shelf off Florida[25]. Wind wave spectral peaks vary, but lie above 0.04 Hz. The model infragravity ocean wave spectrum used in the forcing calculations is also shown.

[1]Lamont Doherty Earth Observatory, Columbia University, Palisades, New York 10964, USA.

coupling of ocean waves with the wave height frequency (σ) and direction (θ) spectrum $f_s(\sigma,\theta)$. To model the forcing of seismic modes by this pressure field, the original Hasselmann expression must be modified with a factor $G(\sigma,H)$ describing the increasing strength of the forcing towards lower frequency and in shallower water (H is water depth) (Fig. 2a). Particle motions beneath ocean waves become increasingly elliptical at shallower depth, with larger horizontal velocities relative to wave height. $G(\sigma,H)$ increases because the coupling depends on the mean squared particle velocity beneath the waves. The calculation of $G(\sigma,H)$ is shown in Supplementary Information. I obtain an expression applicable to the forcing of both Earth seismic modes and microseisms:

$$F_p(\mathbf{k},\omega)\approx\frac{\rho^2 g\omega}{2}G(\omega/2,H)\int_{-\pi}^{\pi}f_s(\omega/2,\theta)f_s(\omega/2,\theta+\pi)\mathrm{d}\theta \quad (1)$$

(here ρ is water density, and g is the local acceleration of gravity). The pressure spectrum has no wavenumber dependence (for small wavenumber), and at a given frequency depends on the ocean wave spectrum at half that frequency: $\sigma=\omega/2$. At typical microseism frequencies (140 mHz) G is equal to 1 (Fig. 2a) except in very shallow water (<40 m depth). At Earth seismic mode frequencies, there is a frequency dependence (ω^{-2}) in G not seen at higher frequency. G at mode frequencies is about 25 dB larger on the continental shelf ($H=30$ m, Fig. 2a) than over ocean basins (>3,000 m). The shallow shelf dominates mode forcing both because the infragravity waves are much larger (an effect amplified by the quadratic dependence of the forcing on the wave spectrum) and because G is larger.

The seismic normal modes appear as narrow spectral lines in the observations (Fig. 1a) because the damping of seismic modes is weak (high quality factor, Q). The ocean wave forcing is best explained as a pressure glut, or jump in pressure acting at the sea surface in a coupled atmosphere–Earth elastic model[16]. The atmospheric component of the ocean wave forcing contributes to a small enhancement of the amplitude of the fundamental mode at 3.7 and 4.4 mHz, but otherwise I ignore weak coupling to the atmosphere and model the forcing as a time-varying vertical point force acting on the Earth's surface. The vertical acceleration spectrum at any site is related to the frequency spectrum of the point force by the function $E(\omega)$ which is the sum of terms describing the resonant forcing of each mode[5]:

$$E(\omega)=\sum_n\sum_l\frac{(2l+1)U_{nl}^4(R)}{4\pi|\Gamma_{nl}(\omega)|^2};$$
$$\Gamma_{nl}(\omega)=\left(\frac{\omega_{nl}}{\omega}\right)^2-\left(1+i\frac{\omega_{nl}}{2Q_{nl}\omega}\right)^2 \quad (2)$$

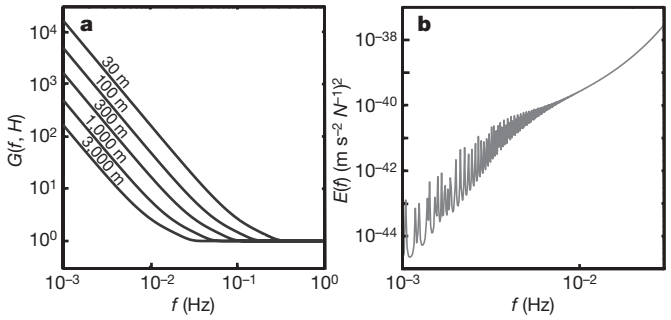


Figure 2 | Increasing mode excitation with frequency under surface forcing is partly balanced by a weakening of the wave interaction mechanism. **a**, The function $G(f,H)$ describing the relative strength of nonlinear wave interaction shown for five water depths H versus frequency f. Note enhancement of wave interaction at shallower water depth and lower frequency. **b**, The function $E(f)$ describing Earth normal mode excitation by a time-varying point vertical force at the Earth's surface versus frequency f. Peaks are associated with mode resonances.

Here ω_{nl} and U_{nl} are respectively the mode resonant frequency and the vertical velocity at the Earth's surface (R) normalized by the mode energy so that $E(\omega)$ describes a balance between mode forcing and dissipation[14]. The mode parameters were calculated using the MINOS program (by F. Gilbert and G. Masters based on ref. 17 applied to the Earth model PA5[18]). The resonant peaks associated with the many modes are obvious in $E(\omega)$ (Fig. 2b). Under forcing by ocean waves, a mode is strongly excited only by those components of the near surface pressure field at frequencies near its resonant frequency and at wavelengths comparable to the spherical harmonic describing that mode. Without a nonlinear mechanism to couple ocean wave energy into high phase velocity components, there would be little coupling to Earth normal modes.

The horizontal scales of the relevant seismic modes are large compared to the widths of the continental shelves, and the forcing is calculated by summing the forcing from many small regions covering the shelves. Beyond the shelf edge, infragravity wave amplitudes rapidly decrease and forcing is negligible. The regions are sufficiently small relative to mode wavelengths to model each region as a temporally fluctuating vertical point force uncorrelated with other regions. With some simplifying assumptions, the predicted vertical acceleration spectrum $A(\omega)$ for the background seismic mode spectrum under wave forcing is

$$A(\omega)=\pi E(\omega)F_p(0,\omega)\Omega_s \quad (3)$$

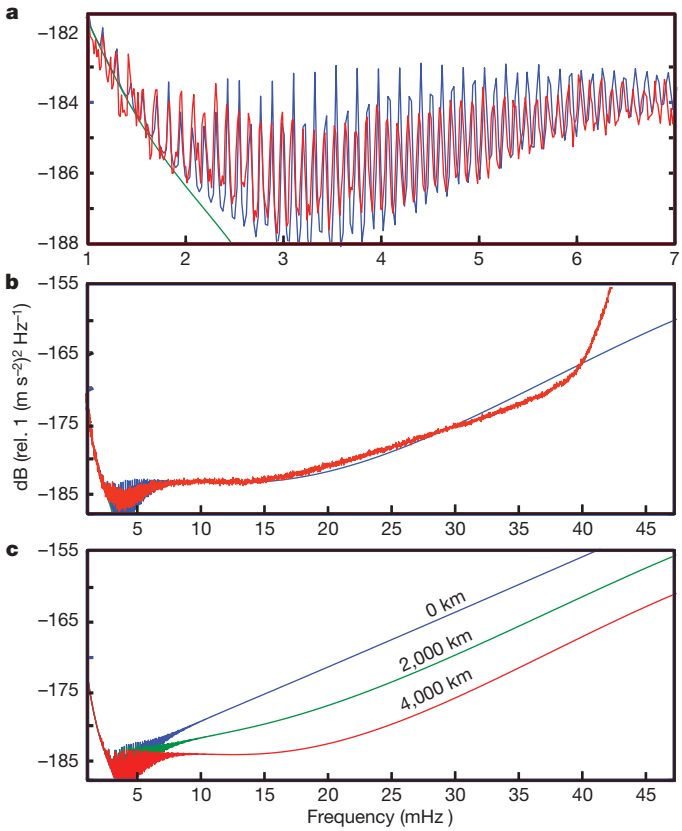


Figure 3 | Comparison of the modelled spectrum with observations. **a**, Spectrum shown (red line: redrawn from ref. 26) is an average of the 32 spectra with the lowest mode energy between 2 mHz and 8 mHz selected from a set of 738,000 hourly spectra derived from 118 GSN stations. This process should select spectra from sites distant from mode sources and time intervals with the least energetic sources. Also shown is a model (blue line) of the excitation of the Earth's hum (at a site 4,000 km into a continent) with an atmospheric gravitational attraction noise model (green line) added. **b**, Same as **a**, but plotted to higher frequency. **c**, Models of the spectrum for sites at three distances from source regions, showing the effect of attenuation.

where Ω_s is the Earth's area covered by shelves (see Supplementary Information).

The predicted spectrum matches observations of the seismic mode background spectrum (Fig. 3a) given a reasonable estimate for the infragravity wave spectrum on the shallow shelf (Fig. 1b). The quadratic dependence of equation (1) on the spectrum ensures that the excitation of normal modes is dominated by regions with the most energetic infragravity waves. A model for the local effect of the time varying gravitational attraction of the atmosphere acting on the seismometer has been added to the predicted spectrum at low frequencies.

Fitting the spectrum above 10 mHz (Fig. 3b, c) requires modifying the model to account for attenuation between the source regions and the seismometers. The model above uses the simplifying assumption that sources are distributed uniformly over the Earth. Sites that are seismically quiet are found within the interior of continents because these sites are remote from ocean waves. The attenuation of ocean wave noise with propagation into the continents is accounted for by adding a factor to $E(\omega)$:

$$E(\omega) = \sum_n \sum_l \frac{(2l+1)U_{nl}^4(R)}{4\pi|\Gamma_{nl}(\omega)|^2} \exp\left[-\frac{\omega X}{u_{nl}Q_{nl}}\right] \qquad (4)$$

Here X represents the distance from nearby ocean noise sources to a site, and u_{nl} is the group velocity associated with a mode (when expressed as a sum of propagating waves). A spectral average from quiet sites is best fitted between 2 and 40 mHz with $X \approx 4{,}000$ km (Fig. 3b). The width of the envelope of $E(\omega)$ and thus the hum spectrum envelope are also controlled by attenuation[19]: the thinner envelope at higher frequency is a result of lower mode Qs at shorter wavelength. The model fits the data within 1.5 dB from 2 mHz to 40 mHz. The remaining differences are equivalent to a 5% difference in mode Qs, and are smaller than the variability between sites, or within spectra from a single site, and less than the biannual cycle in hum energy[4]. The model diverges above 40 mHz because the single frequency microseism peak (Fig. 1a) is generated by a different mechanism (ocean waves interacting with bathymetry[11]).

Previous authors have ascribed the seismic mode background to forcing under atmospheric turbulence[2,4–7]. I believe that this is incorrect, because it was assumed that the pressure signal that can force normal modes is of the same magnitude as the typical pressure fluctuations within atmospheric turbulence: $p \approx \rho U^2$ (ρ, air density; U, wind velocity). This assumption leads to a large overestimation of the forcing because only a tiny component of the turbulent pressure field is associated with the large wavelengths and high phase velocities[20,21] required to excite Earth normal modes. Turbulence can force Earth normal modes in two ways: by developing pressure fluctuations beneath the atmospheric boundary layer that act directly on the Earth's surface, or by coupling first into infrasound above the surface that then propagates downwards to the surface. A strong Mach dependence for these processes ensures that low Mach number turbulence is an inefficient generator of sound[21] or of seismic waves. A model of the pressure spectrum under the atmospheric boundary layer at wavenumbers small enough to drive Earth normal modes calculated from a model for the pressure fluctuations beneath a shear layer[22] predicts levels 150 dB lower than previous papers that supported atmospheric turbulence as the primary source for the Earth's hum (see Supplementary Information). Estimates of the ground forcing under tornadoes[23] and in thunderstorms[24] suggest that these discrete turbulence sources are also insignificant, despite their relatively large Mach numbers.

Careful instrumentation and analysis were required to reveal the presence of a background level of excitation of the seismic normal modes of the Earth[1], and identifying the source as being within the oceans has been equally difficult, requiring processing of data from large seismic arrays[8]. I have shown here that the nonlinear ocean wave interaction mechanism provides the necessary energy to explain the mode background. An alternative mechanism for coupling energy to seismic waves[11] involves the interaction of ocean waves with bathymetry, and this could contribute to mode forcing. Future observations of the temporal correlation between ocean waves and mode spectra should help to constrain the contribution from this alternative mechanism.

Received 2 November; accepted 12 December 2006.

1. Suda, N., Kazunari, K. & Fukao, Y. Earth's background free oscillations. *Science* **279,** 2089–2091 (1998).
2. Nishida, K., Kobayashi, N. & Fukao, Y. Resonant oscillations between the solid Earth and atmosphere. *Science* **287,** 2244–2246 (2000).
3. Ekstrom, G. Time domain analysis of the Earth's background seismic radiation. *J. Geophys. Res.* **106,** 26483–26494 (2001).
4. Kobayashi, N. & Nishida, K. Continuous excitation of planetary free oscillations by atmospheric disturbances. *Nature* **395,** 357–360 (1998).
5. Nishida, K. *et al.* Origin of Earth's ground noise from 2 to 20 mHz. *Geophys. Res. Lett.* **29,** 1413, doi:10.1029/2001GL013862 (2002).
6. Tanimoto, T. Continuous free oscillations: Atmosphere-solid Earth coupling. *Annu. Rev. Earth Planet. Sci.* **29,** 563–584 (2001).
7. Fukao, Y. K. *et al.* A theory of the Earth's background free oscillations. *J. Geophys. Res.* **107** (B9), 2206, doi:10.1029/2001JB000153 (2002).
8. Rhie, J. & Romanowicz, B. Excitation of the Earth's continuous free oscillations by atmosphere-ocean-seafloor coupling. *Nature* **431,** 552–556 (2004).
9. Ekstrom, G. & Ekstrom S.. Correlation of Earth's long-period background seismic radiation with the height of ocean waves. *Eos* 86(52), Fall Meet. Suppl. abstr. S34B–02 (2005).
10. Romanowicz, B., Rhie, J. & Colas, B. Insights into the origin of the Earth's hum and microseisms. *Eos* 86(52), Fall Meet. Suppl. abstr. S31A–0271 (2005).
11. Hasselmann, K. A statistical analysis of the generation of microseisms. *Rev. Geophys.* **1,** 177–209 (1963).
12. Webb, S. C. Broad seismology and noise under the ocean. *Rev. Geophys.* **36,** 105–142 (1998).
13. Dahlen, F. A. & Tromp, J. T. *Theoretical Global Seismology* (Princeton Univ. Press, Princeton, New Jersey, 1998).
14. Herbers, T. H. C. *et al.* Infragravity-frequency (0.005–0.05 Hz) motions on the shelf. Part II: Free waves. *J. Phys. Oceanogr.* **25,** 1063–1079 (1995).
15. Herbers, T. H. C. *et al.* Generation and propagation of infragravity waves. *J. Geophys. Res.* **100** (C12), 24863–24872 (1995).
16. Lognonne, P. *et al.* Computation of seismograms and atmospheric oscillations by normal mode summation for a spherical earth model with a realistic atmosphere. *Geophys. J. Int.* **135,** 388–406 (1998).
17. Woodhouse, J. H. in *Seismological Algorithms, Computational Methods and Computer Programs* (ed. Doornbos, D. J.) 321–370 (Academic, London, 1988).
18. Gaherty, J. B. & Jordan, T. H. Seismic structure of the upper mantle in a central Pacific corridor. *J. Geophys. Res.* **101** (B10), 22291–22309 (1996).
19. Tanimoto, T. The oceanic excitation hypothesis for the continuous oscillations of the Earth. *Geophys. J. Int.* **160,** 276–298 (2005).
20. Meecham, W. C. On aerodynamic infrasound. *J. Appl. Atmos. Terr. Phys.* **33,** 149–155 (1971).
21. Lighthill, J. On sound generated aerodynamically, 1. General theory. *Proc. R. Soc. Lond. A* **211,** 564–587 (1952).
22. Howe, M. S. Surface pressures and sound produced by turbulent flow over smooth and rough walls. *J. Acoust. Soc. Am.* **90,** 1041–1047 (1991).
23. Tatom, F. B. & Vitton, S. J. The transfer of energy from a tornado to the ground. *Seismol. Res. Lett.* **72,** 12–21 (2001).
24. Akhalkatsi, M. *et al.* Infrasound generation by turbulent convection. Preprint at ⟨http://arXiv.org/astro-ph/0409367⟩ (v1, 15 Sept., 2004).
25. Herbers, T. H. C. SAX04 experiment data set ⟨http://www.apl.washington.edu/projects/SAX04/summary.html⟩ (2006).
26. Berger, J. *et al.* Ambient Earth noise: A survey of the Global Seismographic Network. *J. Geophys. Res.* **109,** B11307, doi:10.1029/2004JB003408 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements I thank G. Ekstrom, J. Gaherty and W.W. Webb for discussions, and P. Lognonné and T. Tanimoto for comments and suggestions.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Correspondence and requests for materials should be addressed to the author (scw@ldeo.columbia.edu).

LETTERS

Preserving the evolutionary potential of floras in biodiversity hotspots

Félix Forest[1,2,3]*, Richard Grenyer[3]*, Mathieu Rouget[4], T. Jonathan Davies[5,6], Richard M. Cowling[7], Daniel P. Faith[8], Andrew Balmford[9], John C. Manning[1], Şerban Procheş[10], Michelle van der Bank[11], Gail Reeves[1], Terry A. J. Hedderson[2] & Vincent Savolainen[3]

One of the biggest challenges for conservation biology is to provide conservation planners with ways to prioritize effort. Much attention has been focused on biodiversity hotspots[1]. However, the conservation of evolutionary process is now also acknowledged as a priority in the face of global change[2]. Phylogenetic diversity (PD) is a biodiversity index that measures the length of evolutionary pathways that connect a given set of taxa[3,4]. PD therefore identifies sets of taxa that maximize the accumulation of 'feature diversity'. Recent studies, however, concluded that taxon richness is a good surrogate for PD[5–9]. Here we show taxon richness to be decoupled from PD, using a biome-wide phylogenetic analysis of the flora of an undisputed biodiversity hotspot—the Cape of South Africa. We demonstrate that this decoupling has real-world importance for conservation planning. Finally, using a database of medicinal and economic plant use[10], we demonstrate that PD protection is the best strategy for preserving feature diversity in the Cape. We should be able to use PD to identify those key regions that maximize future options, both for the continuing evolution of life on Earth and for the benefit of society.

The Cape of South Africa is an area of less than 90,000 km^2. Botanically, it is one of the most species-rich areas of the world. There are more than 9,000 plant species, of which about 70% are endemic[11]. For decades it has been noted that a longitudinal gradient in species richness exists across the Cape[12]. The western part, with a predominantly winter rainfall regime, has about twice the density of plant species of the eastern region, which receives rainfall year-round[13]. Higher species richness in the western part has been attributed to variation in speciation and extinction rates as a consequence of differences in historical ecological conditions[14]. In the west, species richness also varies with topography, with the more uniform lowlands having fewer species than the rugged mountain landscapes[13]. Similarly, there are higher numbers of endemic genera in the western part of the Cape[15].

We collected and compiled distribution data for the entire Cape and created an inventory of species and genera per quarter-degree square (QDS; the finest scale available). After extensive fieldwork (2003–2005), we reconstructed the phylogeny of the Cape flora, on the basis of plastid ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (*rbcL*). We used an exemplar from 735 genera, each indigenous to the Cape. Because of computing limitations imposed by the size of the data matrix, phylogenetic relationships were reconstructed using the parsimony ratchet[16]. Molecular branch lengths were optimized using maximum likelihood. Using non-parametric rate smoothing[17], the branch lengths were then transformed to units of absolute time for PD calculation. This is the largest phylogenetic tree yet built for an entire flora.

We compared per-QDS species and genus richness with per-QDS PD (calculated as the length of the subtree that joins the genera in each QDS to the root of the tree[3]). As expected[5–9], we found these diversity indices to be distributed in a similar manner (Fig. 1b, c; linear regressions: PD versus species richness $R^2 = 0.77$, PD versus genus richness $R^2 = 0.96$). These results initially indicated a limited role for PD in conservation planning in this region[6,8,18]. However, this similarity in overall distribution hides key differences in the distribution of these metrics. We found PD to scale with taxon richness, but the scaling to be complex: some regions have more or less PD than would be expected from their taxon richness. Using two tests (a loess regression of per-QDS PD on genus richness, Fig. 1d; and comparing the observed PD in each QDS against an empirical randomization of PD, Fig. 1e) we found a distinctive east–west division in the distribution of PD that broadly corresponds to the climatic zones defined previously[13], with PD for a given number of taxa being higher in the eastern region than in the western.

These results demonstrate that the flora (within QDS) of the western part of the Cape is phylogenetically clustered: it is made up of relatively closely related genera, resulting from multiple radiations over at least the last 25 million years[19–21]. This results in a higher proportion of both shared and short branches, relative to the east, and therefore a lower PD score for a given number of lineages. In contrast, the flora (within QDS) of the eastern region is phylogenetically 'over-dispersed' relative to the western region: it contains genera that are, on average, less closely related to one another. These patterns result from fundamental evolutionary and palaeoclimatic processes in the west[22–24]. Relative over-dispersion in the east is likewise explicable: the eastern flora abuts another biodiversity hotspot (Maputaland–Pondoland–Albany), is highly ecotonal, and contains occasional exemplar genera from unusual ecotypes[11].

We found that these conflicting patterns of taxon diversity and PD invalidate the sole use of taxon richness for conservation actions. Conservation planning is not just about total numbers, but also about marginal gains. To mimic the critical decisions that conservation planners face in the Cape, we set up a series of conservation

[1]South African National Biodiversity Institute, Kirstenbosch Research Centre, Private Bag X7, Claremont 7735, South Africa. [2]Department of Botany, University of Cape Town, University Private Bag, Rondebosch 7700, South Africa. [3]Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3DS, UK. [4]South African National Biodiversity Institute, Biodiversity Planning Unit, Private Bag X101, Pretoria 0001, South Africa. [5]Biology Department, Gilmer Hall, University of Virginia, Charlottesville, Virginia 22904-4328, USA. [6]Institute of Ecology, University of Georgia, Athens, Georgia 30602, USA. [7]Botany Department, Nelson Mandela Metropolitan University, PO Box 77000, Port Elizabeth 6031, South Africa. [8]Australian Museum, 6 College Street, Sydney, New South Wales 2010, Australia. [9]Department of Zoology, University of Cambridge, Downing Street, Cambridge, CB2 3EJ, UK. [10]Centre for Invasion Biology, Private Bag X1, University of Stellenbosch, Matieland 7602, South Africa. [11]Department of Botany and Plant Biotechnology, University of Johannesburg, PO Box 524, Auckland Park 2006, South Africa.
*These authors contributed equally to this work.

scenarios based upon complementarity. In each scenario, an additional locality that maximizes gain in a biodiversity index is to be included in an existing partial set of conservation areas. We initially chose additions to partial sets based upon taxon richness, and examined the marginal gain in PD experienced. We then contrasted these gains by choosing additions to partial sets based directly on PD. The results show that gains in taxon richness and gains in PD are decoupled (Fig. 2 and Supplementary Information). Typically, selection for conventional taxon complementarity misses localities that would provide large gains in PD.

Why does it matter that PD is not well captured by conventional taxon-based policies? We argue that maximizing PD is the best bet-hedging strategy. By maximizing feature diversity we maximize option value: the possibility of having the right feature at hand in an uncertain future. We use a practical example to illustrate this point. We identified all genera in the Cape with species of medicinal or economic importance[10]. We divided these genera into three types of use (food, medicine and other). Using a randomization test, we found that each type of use is phylogenetically clumped (all $P < 0.01$) and that different categories of use are clustered in different parts of the phylogeny (Supplementary Information). So how should we have designed a conservation strategy to preserve useful plants if this distribution were not known? Simply choosing samples of the largest possible number of genera permits the selection of a set of genera that are themselves phylogenetically clumped: such a set might include many genera of one type of use but this must come at a cost to the

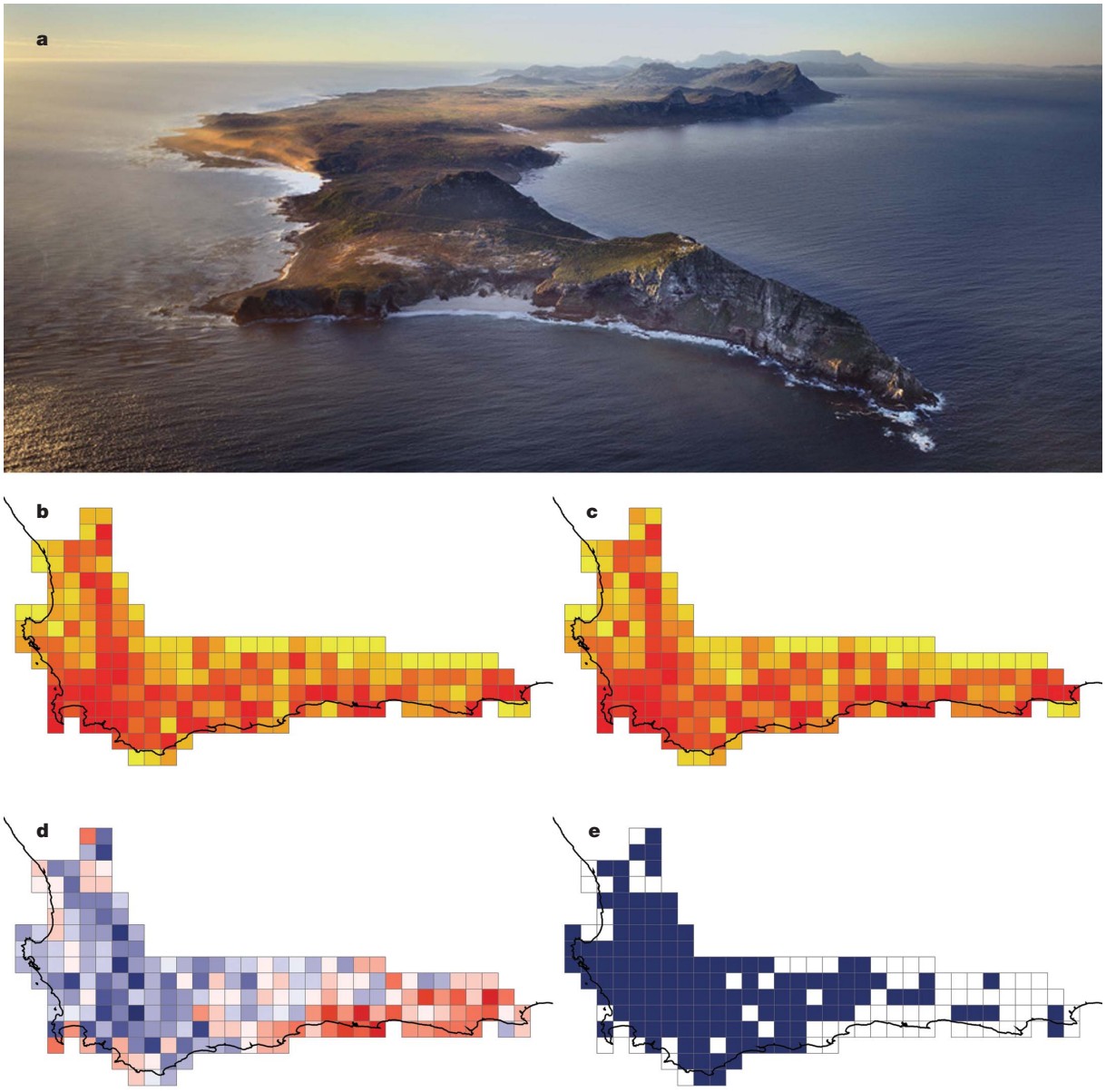


Figure 1 | Taxon richness and phylogenetic diversity in the Cape. **a**, This biodiversity hotspot, which includes the renowned Cape Peninsula, Cape of Good Hope and Table Mountain (in the distance), dominated by fynbos vegetation, is situated at the southern tip of Africa. Picture credit: A. Proust/iAfrika. **b**–**e**, Maps of 201 QDS covering the entire Cape region. **b**, Genus richness (ten quantile intervals from yellow to deep red). **c**, PD calculated using NPRS absolute age estimates in million years (colour code as for **b**). **d**, Residuals from a loess regression of PD (calculated using NPRS absolute age estimates) on genus richness. QDS with negative residuals are indicated in blue, and those with positive residuals are shown in red (shading increments of half a standard deviation). **e**, The spatial distribution of unusual PD values, as assessed by comparing the observed PD in each QDS with 10,000 PD values calculated by random selection of the same number of genera from the Cape flora. Cells with significantly lower PD ($P < 0.05$, two-tailed) than expected are shaded in blue. A similar pattern was found when the tree was simplified to reflect the phylogeny-based taxonomy of the Angiosperm Phylogeny Group[30] (Supplementary Information).

other two types, because they are found in different parts of the phylogeny. Choosing a selection of over-dispersed taxa based on PD would have maximized the probability of having representatives of each of the three classes of use. In fact, we found that a set of cells chosen to maximize PD complementarity (blue line, Fig. 2) samples all useful genera in 13 QDS, while a set chosen to maximize taxon complimentarity (black line, Fig. 2) requires 15 QDS to do so, although the majority of useful genera in both cases are sampled in the first few QDS. In an uncertain future, where we are not yet sure of the sort of plant features we will need, we argue that incorporating gains in PD into conservation planning is the best strategy.

It is not just our own options, though, that we need to keep open. We do not know the characteristics that species in the Cape will need to adapt and diversify in a future of climatic change. We therefore argue that maximizing PD will in turn maximize the options for future diversification. The many radiations in the western part of the Cape may well be a reason[25] to see the region as one of high evolutionary potential. However, although it is possible, we see no reason why future speciation regimes must be the same as those that gave rise to the historical diversification in the western part of the Cape. Throughout the history of angiosperms, diversification has been a complex process in which the propensity to diversify was highly labile and dependent upon many different traits at different times[26]. Our recommendation would not be to reject recently diversifying sites in the west as conservation targets, but to ensure that PD is maximized by inclusion of suitable areas in the east into existing conservation schemes. Balancing these two diversity indices is now at least an algorithmic problem for which we have suitable tools[27]. We also note that scale is important. For example, our phylogenetic tree does not include lineages that are not found in the Cape, but we have chosen the most biologically sound limits: a phytogeographic delimitation[28] that falls within a single country and can therefore be managed under a single coordinated conservation response[29]. Any conservation plan that operates at less than a global scale will always be at risk of finding solutions that are optimal only within the region being considered.

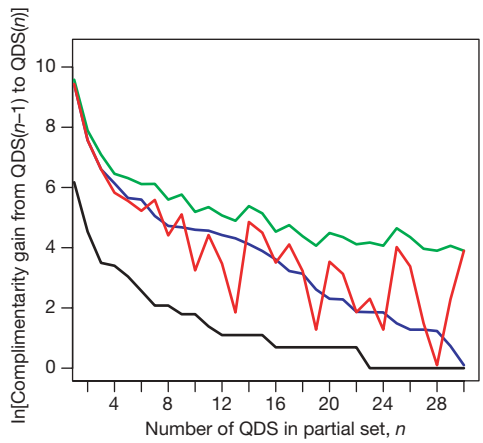


Figure 2 | Complementarity analysis of PD and genus richness. A series of conservation scenarios based on complementarity were set up with a simple greedy algorithm: for each partial set defined along the *x* axis, we identified the additional QDS that provided the highest possible genus-based complementarity (black). We then calculated the PD complementarity that would be provided by this same QDS (red), as well as the alternative additional QDS that would provide the highest possible PD complementarity (green). Whereas comparisons of diversity measures have usually focused on richness, it is now apparent that decision-making depends on marginal gains (complementarity values) and these must be the basis for comparisons. Here we show that gains in genus richness are poor predictors of gains in PD (contrast red and green lines). We also show the independent series of QDS that provide the highest PD-based complementarity (blue).

We have shown that a simple correspondence between taxon richness and PD can hide a fundamental decoupling of biodiversity indices, with drastically different conservation outcomes if only one of the indices is used. The Cape is one of the most well-studied hotspots, so our findings clearly raise the possibility that similar decoupling may be found in others. Further, we know from simulations[7] that a decoupling of PD and taxon richness is most likely when the underlying phylogeny is unbalanced and there is strong phylogeographic structure; both these are epiphenomena of endemic radiations. We conclude by suggesting that because biodiversity hotspots are defined in part by their richness in endemics[1], they are precisely the areas in which a decoupling of PD and taxon richness is most likely—as is observed here.

METHODS

DNA sequencing and phylogeny reconstruction. We sampled one exemplar species for 735 of the 943 genera of angiosperms currently recognized in the Cape (~78%) and obtained sequence data for the plastid *rbcL* exon (ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit). Phylogenetic relationships were reconstructed using the parsimony ratchet[16] method with 15% of the characters perturbed and 200 iterations; ten independent parsimony ratchet searches were performed and the shortest trees resulting from these independent searches were used to create a consensus tree. Clade support was assessed with 500 bootstrap replicates. One of the most parsimonious trees from the parsimony ratchet analysis was chosen as a best hypothesis of relationships for the Cape plant genera. PD calculations were performed using branch lengths (maximum parsimony and maximum likelihood) and age estimates (non-parametric rate smoothing[17]; relative time divergences were transformed into absolute ages using twelve well-characterized fossils; see Supplementary Information).

Distribution of phylogenetic diversity. The distribution of genera within the Cape was compiled as a binary matrix of absence/presence per quarter degree square (QDS; approximately 25 km × 27 km) using data from the Pretoria National Herbarium database (PRECIS). The spatial pattern of the relationship between PD and taxon richness was revealed by plotting the residuals for a loess regression of per-QDS PD on taxon richness. To locate QDS with significantly higher or lower PD than expected from their taxon richness, the PD in each QDS was compared with 10,000 PD values for sets of genera of the same size, sampled without replacement (Supplementary Information).

Medicinal and economic species. A randomization procedure was used to assess whether the distribution of medicinal and economic species is constrained by the phylogeny or randomly distributed across lineages. To be considered of medicinal and/or economic use, a given genus must have at least one species found in the Cape that is recorded in the database of the Survey of Economic Plants for Arid and Semi-Arid Lands (SEPASAL[10]; Supplementary Information).

Received 27 October 2006; accepted 9 January 2007.

1. Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403,** 853–858 (2000).
2. Mace, G. M., Gittleman, J. L. & Purvis, A. Preserving the tree of life. *Science* **300,** 1707–1709 (2003).
3. Faith, D. P. Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61,** 1–10 (1992).
4. Vane-Wright, R. I., Humphries, C. J. & Williams, P. H. What to protect? Systematics and the agony of choice. *Biol. Conserv.* **55,** 235–254 (1991).
5. Brooks, T. M. *et al.* Global biodiversity conservation priorities. *Science* **313,** 58–61 (2006).
6. Polasky, S., Csuti, B., Vossler, C. A. & Meyers, S. M. A comparison of taxonomic distinctness versus richness as criteria for setting conservation priorities for North American birds. *Biol. Conserv.* **97,** 99–105 (2001).
7. Rodrigues, A. S. L., Brooks, T. M. & Gaston, K. J. in *Phylogeny and Conservation* (eds Purvis, A., Gittleman, J. L. & Brooks, T. M.) 101–119 (Cambridge University Press, Cambridge, UK, 2005).
8. Rodrigues, A. S. L. & Gaston, K. J. Maximising phylogenetic diversity in the selection of networks of conservation areas. *Biol. Conserv.* **105,** 103–111 (2002).
9. Torres, N. M. & Diniz, J. A. F. Phylogenetic autocorrelation and evolutionary diversity of Carnivora (Mammalia) in conservation units of the New World. *Genet. Mol. Biol.* **27,** 511–516 (2004).
10. Survey of Economic Plants for Arid and Semi-Arid Lands. *(SEPASAL)* database ⟨http://www.rbgkew.org.uk/ceb/sepasal/internet/⟩ (Royal Botanic Gardens, Kew, London, 1999).
11. Goldblatt, P. & Manning, J. C. Plant diversity of the Cape region of southern Africa. *Ann. Missouri Bot. Gard.* **89,** 281–302 (2002).
12. Linder, H. P. The radiation of the Cape flora, southern Africa. *Biol. Rev.* **78,** 597–638 (2003).

13. Cowling, R. M. & Lombard, A. T. Heterogeneity, speciation/extinction history and climate: explaining regional plant diversity patterns in the Cape floristic region. *Divers. Distrib.* **8,** 163–179 (2002).
14. Cowling, R. M., Cartwright, C. R., Parkington, J. E. & Allsopp, J. C. Fossil wood charcoal assemblages from Elands Bay Cave, South Africa: implications for Late Quaternary vegetation and climates in the winter-rainfall fynbos biome. *J. Biogeogr.* **26,** 367–378 (1999).
15. Linder, H. P. & Midgley, J. J. Taxonomy, compositional biodiversity and functional biodiversity of fynbos. *S. Afr. J. Sci.* **90,** 329–333 (1994).
16. Nixon, K. C. The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics* **15,** 407–414 (1999).
17. Sanderson, M. J. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.* **14,** 1218–1231 (1997).
18. Williams, P. H. & Humphries, C. J. in *Biodiversity: a Biology of Numbers and Difference* (ed. Gaston, K. J.) (Blackwell Science, Oxford, UK, 1996).
19. Goldblatt, P. *et al.* Radiation in the Cape flora and the phylogeny of peocock irises *Moraea* (Iridaceae) based on four plastid DNA regions. *Mol. Phyl. Evol.* **25,** 341–360 (2002).
20. Linder, H. P. Evolution of diversity: the Cape flora. *Trends Plant Sci.* **10,** 536–541 (2005).
21. Richardson, J. E. *et al.* Rapid and recent origin of species richness in the Cape flora of South Africa. *Nature* **412,** 181–183 (2001).
22. Cowling, R. M. & Proches, S. in *Plant Diversity and Complexity Patterns: Local, Regional and Global Dimensions* (eds Friis, I. & Balslev, H.) 273–288 (The Royal Danish Academy of Sciences And Letters, Copenhagen, 2005).
23. Midgley, G. F., Hannah, L., Roberts, R., MacDonald, D. J. & Allsopp, J. Have Pleistocene climatic cycles influenced species richness pattern in the greater Cape Mediterranean Region? *J. Mediterr. Ecol.* **2,** 137–144 (2001).
24. Proches, S., Wilson, J. R. U. & Cowling, R. M. How much evolutionary history in a 10x10 m plot? *Proc. R. Soc. Lond. B* **273,** 1143–1148 (2006).
25. Erwin, T. L. An evolutionary basis for conservation strategies. *Science* **253,** 750–752 (1991).
26. Davies, T. J. *et al.* Darwin's abominable mystery: insight from a supertree of the angiosperms. *Proc. Natl Acad. Sci. USA* **101,** 1904–1909 (2004).
27. Wilson, K. A., McBride, M. F., Bode, M. & Possingham, H. P. Prioritizing global conservation efforts. *Nature* **440,** 337–340 (2006).
28. Goldblatt, P. & Manning, J. C. *Cape Plants, a Conspectus of the Cape Flora in South Africa* (National Botanical Institute of South Africa, Cape Town, South Africa, 2000).
29. Mooers, A. O., Heard, S. B. & Chrostowski, E. in *Phylogeny and Conservation* (eds Purvis, A., Brooks, T. L. & Gittleman, J. L.) (Oxford Univ. Press, Oxford, 2005).
30. APG. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Bot. J. Linn. Soc.* **141,** 399–436 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Arnold, K. Balele, W. Barrington, N. Bergh, F. Conrad, L. Csiba, C. Cupido, A. Dold, the Fourcade Botanical Club, K. Davis, J. Donaldson, P. Drew, T. Fulcher, G. Gardiner, J. Gittleman, P. Goldblatt, N. Helme, E. Kapinos, A. Khunou, N. B. Lester, A. Mabunda, M. Powell, D. Snijman, K. Tolley, T. Trinder-Smith, A. G. Verboom, E. van Jaarsveld, S. Vetter, C. Williams, M. Wolfson, F. Woodvine, and especially I. Nänni, for assistance; the conservation authorities of the Western, Eastern and Northern Cape in South Africa for granting collecting permits as well as the managers of nature reserves and private landowners; A. Proust/iAfrika for the picture in Fig. 1; and T. Barraclough, M. Chase and H. Possingham for comments on the manuscript. We thank the Darwin Initiative for the Survival of Species, the South African National Biodiversity Institute, the University of Cape Town, the Royal Botanic Gardens Kew, the Bentham-Moxon Trust, the US National Science Foundation, the University of Virginia and the European Commission (HOTSPOTS/EDIT) for funding.

Author Information DNA sequences have been deposited at GenBank/EMBL under accession numbers AM234779–AM235167 (see also Supplementary Information). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to F.F. (f.forest@kew.org) and V.S. (v.savolainen@kew.org).

LETTERS

Promotion of Hras-induced squamous carcinomas by a polymorphic variant of the *Patched* gene in FVB mice

Yuichi Wakabayashi[1], Jian-Hua Mao[1], Ken Brown[2]†, Michael Girardi[3] & Allan Balmain[1,4]

Mice of the C57BL/6 strain are resistant to the development of skin squamous carcinomas (SCCs) induced by an activated Ras oncogene, whereas FVB/N mice are highly susceptible[1]. The genetic basis of this difference in phenotype is unknown. Here we show that susceptibility to SCC is under the control of a carboxy-terminal polymorphism in the mouse *Ptch* gene. F_1 hybrids between C57BL/6 and FVB/N strains ((B6FVB)F_1) are resistant to Ras-induced SCCs, but resistance can be overcome either by elimination of the C57BL/6 *Ptch* allele ($Ptch^{B6}$) or by overexpression of the FVB/N *Ptch* allele ($Ptch^{FVB}$) in the epidermis of *K5Hras*-transgenic (B6FVB)F_1 hybrid mice. The human *Patched* (*PTCH*) gene is a classical tumour suppressor gene for basal cell carcinomas and medulloblastomas, the loss of which causes increased signalling through the *Sonic Hedgehog* (*SHH*) pathway[2–5]. SCCs that develop in $Ptch^{B6+/-}$ mice do not lose the wild-type *Ptch* gene or show evidence of increased SHH signalling. Although $Ptch^{FVB}$ overexpression can promote SCC formation, continued expression is not required for tumour maintenance, suggesting a role at an early stage of tumour cell lineage commitment. The *Ptch* polymorphism affects Hras-induced apoptosis, and binding to Tid1, the mouse homologue of the *Drosophila l(2)tid* tumour suppressor gene. We propose that *Ptch* occupies a critical niche in determining basal or squamous cell lineage, and that both tumour types can arise from the same target cell depending on carcinogen exposure and host genetic background.

About 1.5% of the *K5Hras* transgene-positive animals on the mixed (C57BL/6 × CBA) background[6] developed 'early-onset' carcinomas within the first few weeks after birth (Fig. 1a). Figure 1a shows the average phenotype distribution in these mice after several rounds of backcrossing to the strains indicated. Breeding with the FVB/N strain consistently produced litters that showed about 30% tumour-positive mice within the first 3 weeks after birth, and a substantial proportion of the remaining transgene-positive FVB/N mice went on to develop keratoacanthomas or carcinomas after weaning. The C57BL/6 mice carrying the *K5Hras* transgene failed to develop any early-onset carcinomas, although about 20% of the adults developed keratoacanthoma, as well as sebaceous adenomas. The resistance conferred by the C57BL/6 background was dominant over susceptibility, because F_1 hybrids between the C57BL/6 and FVB/N strains ((B6FVB)F_1) failed to show any evidence of carcinoma formation (Supplementary Table 1, and data not shown). The lack of tumour development in the C57BL/6 background was not due to strain-specific silencing of the transgene, because expression analysis demonstrated that transgene mRNA levels were not affected by genetic background (Supplementary Fig. 1a).

Linkage analysis on about 440 backcross mice generated by crossing the (B6FVB)F_1 hybrids with the susceptible FVB/N parental strain showed that one major locus was responsible for the resistance phenotype, located on mouse chromosome 13 (Fig. 1b). The 95% confidence interval was about 1.4 megabases (Mb), and contained two major candidate genes: *Patched* and the *Fanconi's Anaemia* C gene (Fig. 1c). These genes are separated by an interval of only 100 kilobases, and the linkage could not be further refined by the use of additional genetic markers. Both genes have been implicated in responses to DNA damage and apoptosis, as well as in cancer[2,3,7].

We initially focused on *Ptch* because of its known involvement in susceptibility to skin cancer. We sequenced the *Ptch* exons to determine whether any polymorphisms exist between C57BL/6 and FVB/N strains that could explain the observed results. A polymorphism was detected in the *Ptch* C terminus (T1267N; Fig. 1d), within a conserved region of the cytoplasmic domain. To test the functional involvement of this allele in squamous carcinoma resistance, we obtained C57BL/6 mice carrying inactivated alleles of *Ptch* and generated *K5Hras*/$Ptch^{+/-}$ (B6FVB)F_1 mice. Elimination of the germline C57BL/6 allele of *Ptch* in these *K5Hras*/$Ptch^{+/-}$ (B6FVB)F_1 hybrid mice led to squamous carcinoma development in about 30% of the offspring (Fig. 2a, and Supplementary Table 1). We conclude that the C57BL/6 allele carrying the threonine at position 1267 is responsible for the resistance to SCC development in these transgenic animals, and that inactivation of this *Ptch* allele in the germline leads to carcinoma formation.

Mutations of *Ptch* in both mouse and human skin cancers are accompanied by loss of the remaining wild-type allele and increased SHH signalling[5,8,9], leading to the formation of a basal cell carcinoma (BCC). Formation of SCCs is not normally associated with stimulation of the SHH signalling pathway. Tumours from the *K5Hras*/$Ptch^{+/-}$ mice were clearly SCCs with differing degrees of keratinization (Fig. 2a), but in contrast to the situation observed with BCCs from $Ptch^{+/-}$ mice[8], all retained the wild-type $Ptch^{FVB}$ allele (Fig. 2b). We examined other components of the SHH signalling pathway in normal skin, in skin induced to proliferate by treatment with 12-*O*-tetradecanoylphorbol-13-acetate (TPA), and in SCC tumours induced by the *K5Hras* transgene in mice heterozygous ($Ptch^{+/-}$) or homozygous for the $Ptch^{FVB}$ allele. Although *Ptch* and its downstream effector *Gli2* are expressed in normal skin from newborn and adult mice (Fig. 2c, d), neither of these SHH pathway components seems to be expressed in SCCs from the same animals. In addition, induction of proliferation and differentiation by treatment with TPA is associated with reduced levels of signalling through the SHH pathway (Supplementary Fig. 1b).

To test the possibility that the C57BL/6 *Ptch* allele might act as a dominant negative over the FVB/N allele to confer tumour resistance, we generated transgenic mice that express $Ptch^{FVB}$ in the epidermis under the control of the keratin 14 (K14) promoter[10].

[1]Cancer Research Institute, University of California at San Francisco, 2340 Sutter Street, San Francisco, California 94143, USA. [2]CRC Department of Medical Oncology, University of Glasgow, Garscube Estate, Switchback Road, Glasgow G61 1BD, UK. [3]Department of Dermatology, Yale University School of Medicine, New Haven, Connecticut 06520, USA. [4]Department of Biochemistry and Biophysics, University of California at San Francisco, 2340 Sutter Street, San Francisco, California 94143. †Present address: CXR Biosciences Ltd, James Lindsay Place, Dundee Technopole, Dundee DD1 5JJ, UK.

Although expression of the $Ptch^{FVB}$ transgene was detected at variable levels in the different transgenic lines (Fig. 3a), the mice had no obvious phenotype, and the skin had a histologically normal appearance (data not shown). Nevertheless, when crossed with the tumour-resistant *K5Hras* (B6FVB)F_1 mice, a proportion of the animals born developed clear early-onset SCCs (Fig. 3d, e, and Supplementary Table 1). A total of 250 mice were studied from these crosses, 59 of which were double transgenics carrying both *K5Hras* and $K14Ptch^{FVB}$ transgenes. Of these double-transgenic mice, 16 developed histologically verified SCCs (27%), in contrast with about 30% of the B6 $Ptch^{+/-}$ mice in the (B6FVB)F_1 background (Supplementary Table 1). These data are highly significant ($P = 3.34 \times 10^{-10}$), because no early-onset tumours were seen in 152 single-transgenic *K5Hras* mice on the equivalent F_1 background.

Tumours that arise as a consequence of transgenic overexpression of a particular oncogene are normally dependent on the continued

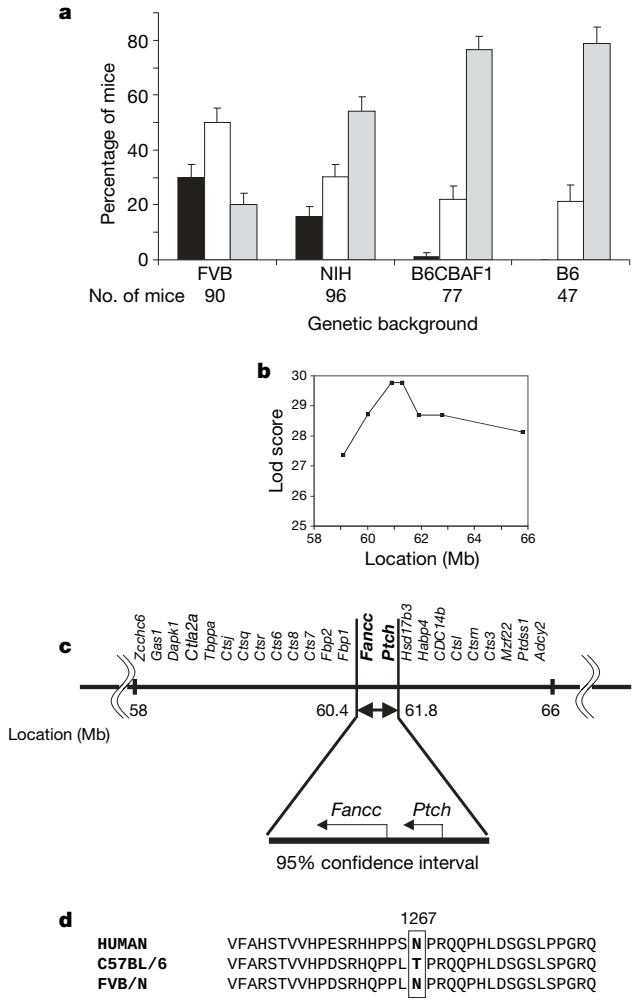


Figure 1 | Genetic linkage of the C57BL/6 SCC resistance to a C-terminal polymorphism in the *Ptch* gene. **a**, *K5Hras* carcinoma phenotype is dependent on genetic background. Black columns, early onset; white columns, late onset; grey columns, no phenotype. The error bars show the standard deviation based on a polynomial distribution for the given numbers of mice of each strain. **b**, Lod scores of the linkage region on chromosome 13, using a total of 440 *K5Hras* F_1 backcross mice. The 95% confidence interval is about 1.4 Mb. **c**, The physical map of chromosome 13 around the candidate interval. The map was constructed with data from the Celera database (http://www.celera.com) and the Ensembl database (http://www.ensembl.org). Two genes, *Fancc* and *Ptch*, lie within the 95% confidence interval. **d**, The partial amino acid sequences of the cytoplasmic domain of PTCH (Ptch) protein of human and of C57BL/6 and FVB/N mice. There is one amino acid polymorphism (T1267N) between C57BL/6 and FVB/N.

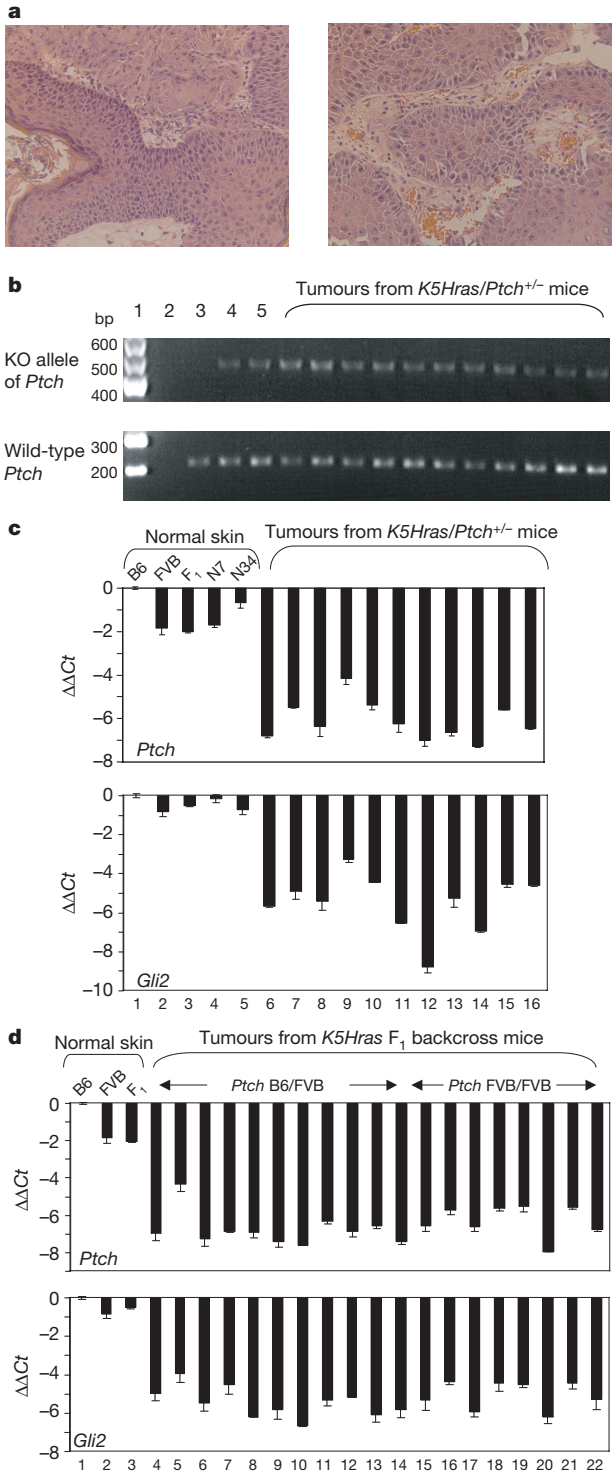


Figure 2 | Expression of *Ptch* and *Gli2* in tumours of F_1 backcross and *K5Hras/*$Ptch^{+/-}$ mice. **a**, Histological appearance of squamous carcinomas from two different ten-day-old *K5Hras/*$Ptch^{+/-}$ mice. Original magnification ×200; haematoxylin/eosin staining. **b**, Retention of the wild-type *Ptch* allele in tumours from *K5Hras/*$Ptch^{+/-}$ mice. Lane 1, 100-bp ladder; lane 2, negative control (no DNA); lane 3, normal skin from wild-type (B6FVB) F_2; lanes 4 and5: normal skin from *K5Hras/*$Ptch^{+/-}$ mice; the remaining lanes contained tumour DNAs from *K5Hras/*$Ptch^{+/-}$ mice. **c**, **d**, Quantitative TaqMan analysis of *Ptch* and *Gli2* expression in normal skin and tumour samples. **c**, B6, FVB and F_1 represent complementary DNAs from normal skins, whereas N7 and N34 are two independent *K5Hras/*$Ptch^{+/-}$ mice. **d**, cDNAs from tumours of *K5Hras* F_1 backcross mice. *Ptch* B6/FVB indicates that the *Ptch* locus is heterozygous, whereas FVB/FVB indicates FVB/N homozygotes. The control sample is B6 RNA in lane 1. $\Delta\Delta Ct$ values were calculated as described in Methods. All tumours showed strongly decreased signals for *Ptch* and *Gli2* expression compared with normal skin samples. Error bars indicate s.d. ($n = 3$).

expression of that gene for the maintenance of tumour growth[11,12]. Although the *K14Ptch*^{FVB}transgene was responsible for tumour induction in the double-transgenic mice, expression of both transgene and the endogenous *Ptch* gene(s) was switched off in the SCCs (Fig. 3b, c). A similar loss of endogenous *Ptch* expression was seen all other SCCs examined, regardless of genetic background (Fig. 2d). In contrast, *K5Hras* expression was clearly maintained in the same tumours and was, if anything, overexpressed compared with the corresponding normal uninvolved skin (Fig. 3b). These data suggest that although the *Ptch*^{FVB} allele may be required for the earliest stages of SCC development, either it is not required or it is positively silenced during the tumour growth phase.

Several studies have implicated *Ptch* in functions that may have a role in cell fate decisions, apoptosis or cell cycle progression[13–17]. We tested the possibility that the two alternative alleles of *Ptch* may differ in their capacity to affect apoptosis. The two alleles were cloned into expression vectors and transfected into NIH 3T3 cells together with an activated *HRAS* oncogene. Transfection of mutant *HRAS* induced apoptosis in a proportion of the NIH 3T3 cells, and this effect was in part rescued by co-transfection of the *Ptch* expression constructs. Consistently, the extent of rescue was greater for the *Ptch*^{FVB} allele than for the C57BL/6 allele (Fig. 4a), despite equivalent levels of expression of the two alleles in recipient cells (Fig. 4b). A similar experiment was also performed with C5N immortalized keratinocytes rather than NIH 3T3 cells, with almost identical results (Fig. 4c, d). This provides a potential mechanism for the susceptibility to Hras-induced squamous tumour development shown by FVB/N mice[1]: preferential rescue of the Hras-initiated cells by the FVB/N Ptch protein.

The mouse *Tid1* gene is the homologue of the *Drosophila lethal (2) tumorous imaginal disks* (*l(2)tid*) tumour suppressor gene[18], and was identified as a C-terminal binding partner of mouse Ptch[19]. Additional studies demonstrated that *Tid1* is involved in apoptosis and senescence signalling through many different routes, including the nuclear factor (NF)-κB, Ras and interferon signalling pathways[20–23]. Figure 4e shows that Ptch^{B6} preferentially forms a complex with Tid1 in comparison with the corresponding Ptch^{FVB}, as shown by immunoprecipitation of endogenous Tid1 (refs 19, 20) or transfected haemagglutinin (HA)-tagged Ptch from 293T cells. Although the existence of two Tid1 isoforms with different activities has been documented[18,22], Ptch^{B6} seemed to show preferential binding to the smaller Tid1-S isoform that has been linked to Ras-induced apoptosis and senescence[22]. These data therefore establish a biochemical difference between the Ptch^{B6} and Ptch^{FVB} proteins and link the polymorphism to differential signalling through the Tid1 tumour suppressor pathway.

The observation of a single major locus that controls carcinoma susceptibility in this K5Hras model contrasts with previous data

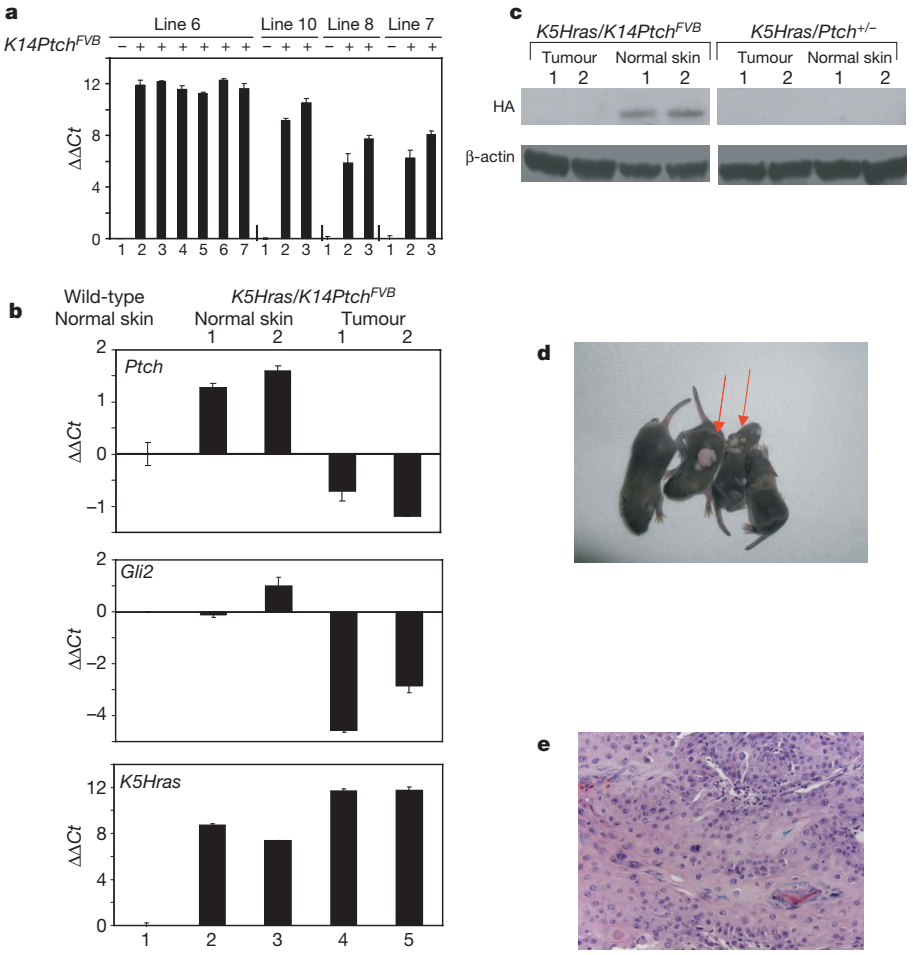


Figure 3 | Analysis of *K14Ptch*^{FVB} transgenic mice. **a**, TaqMan analysis of *K14Ptch*^{FVB} transgene-specific RNA expression level in normal skin from different lines. **b**, TaqMan analysis of total *Ptch*, *Gli2*, and *K5Hras*. The *K14Ptch*^{FVB} transgene is expressed in uninvolved skin, as shown by the higher level of total *Ptch* RNA than in wild-type mice. In the two tumours, both the transgene and endogenous *Ptch* and the endogenous *Gli2* are downregulated in comparison with normal skin. *K5Hras*-specific expression is detected in the uninvolved transgenic skin and is relatively increased in the tumours. **c**, Western blotting analysis of *K14Ptch*^{FVB} transgene in normal skin and tumour. Transgene expression is detected by anti-HA antibodies. **d**, Gross appearance of *K5Hras/K14Ptch*^{FVB} double-transgenic mice ten days after birth. The three mice on the left are the double-transgenic mice, two of which developed SCCs on the back. Arrows indicate tumours. **e**, Histological verification of SCC from ten-day-old *K5Hras/K14Ptch*^{FVB} double-transgenic mice. Error bars indicate s.d. ($n = 3$).

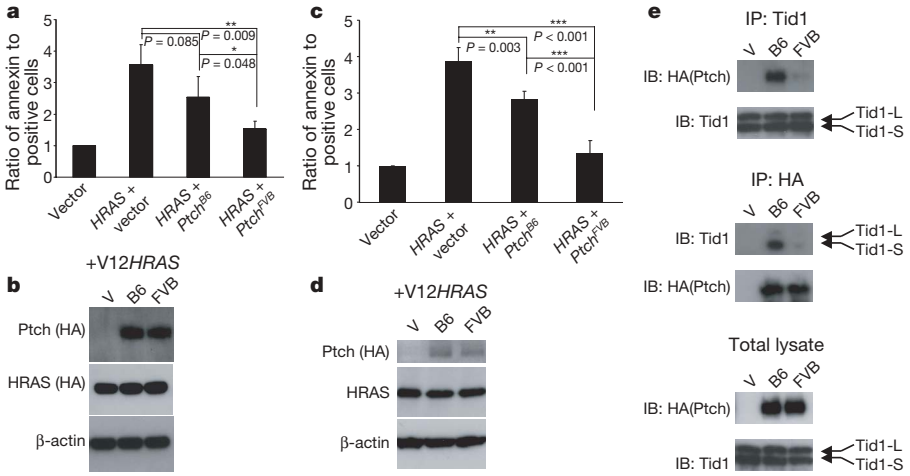


Figure 4 | Effect of Ptch on HRAS-induced apoptosis and differential binding of Ptch^{B6} and Ptch^{FVB} to Tid1. **a**, pcDNA3.1 *Ptch^{B6}* and *Ptch^{FVB}* were transiently transfected into NIH 3T3 cells together with pcDNA3.1 V12*HRAS*. The Ptch^{B6} has a slight but statistically insignificant effect on apoptosis induced by HRAS, whereas the Ptch^{FVB} significantly reduces apoptosis. **b**, Western blotting analysis of expression level of Ptch^{FVB}, Ptch^{B6} and V12Hras in transfected NIH 3T3 cells. **c**, C5N keratinocytes were infected with MSCV retrovirus expressing *Ptch^{B6}* or *Ptch^{FVB}*, followed by infection with pLXSP3 retrovirus expressing V12*HRAS*. The Ptch^{B6} has a weaker but statistically significant effect on apoptosis induced by HRAS, whereas Ptch^{FVB} protein more significantly reduces apoptosis levels. **d**, Western blotting analysis of expression level of Ptch^{FVB}, Ptch^{B6} and V12Hras in transfected C5N keratinocytes. **e**, 293T cells were transfected with HA-tagged Ptch^{B6} or Ptch^{FVB} and immunoprecipitated (IP) with anti-Tid1 or anti-HA antibody. Tid1-L denotes the longer isoform of Tid1 protein, and Tid1-S is the shorter isoform. Ptch^{B6} binds Tid1 much more strongly than Ptch^{FVB}. IB, immunoblotting. Error bars indicate s.d. ($n = 3$).

showing that many low-penetrance tumour-susceptibility loci control multistage carcinogenesis of the skin and other tissues[24,25]. This is probably because the K5Hras model leads to carcinoma development shortly after birth, as a result of expression of the *Hras* transgene in the cells of the epidermal stem niche in the bulge region of the hair follicle[6,26]. This acute phenotype shows substantially less genetic heterogeneity than standard models, enabling rapid high-resolution mapping of the major locus involved. Our data do not allow us to reach the conclusion that *Ptch* is the only gene in the region that affects this phenotype. However, the susceptibility phenotype can be replicated to about the same extent, either by knocking out the *Ptch^{B6}* allele or by overexpressing the *Ptch^{FVB}* allele. If the *Fancc^{B6}* allele (or another neighbouring gene) is involved in conferring resistance, its effects must be relatively minor and are outweighed by the expression of the *Ptch^{FVB}* allele in the *K14Ptch* transgenic mice.

The model that we believe best explains our results is presented in Supplementary Fig. 2a, b. The observation that increased expression of the *Ptch^{FVB}* allele can promote Hras-induced SCCs, but continued expression of the transgene is not required for tumour growth, suggests that *Ptch* contributes positively to a very early stage of carcinoma formation, possibly to the cancer stem-cell fate decision (Supplementary Fig. 2b). Although loss of *Ptch* and activation of SHH signalling are undoubtedly critical for BCC formation, the opposite may be true for SCC formation.

This model may mimic the mechanisms of stem-cell fate decisions that are operative during normal development. The activated *Hras* oncogene that is responsible for initiation of chemical carcinogenesis[27] and for the phenotype of the *K5Hras* mice[6] is effectively leading to constitutive activation of the Raf–mitogen-activated protein kinase (MAPK)–extracellular signal-regulated kinase (ERK) signalling pathway that is physiologically stimulated by endogenous levels of growth factors such as epidermal growth factor, transforming growth factor-α or fibroblast growth factor. Studies of fibroblasts *in vitro* have demonstrated that whereas some combinations of secreted factors may lead to cell growth, changes in relative levels of the same factors can lead to the activation of cell death pathways[28]. It is likely that similar threshold levels of signalling through developmentally important factors such as SHH, Wnt or epidermal growth factor can influence stem-cell fate decisions by controlling entry into apoptosis or senescence[16,17]. These signalling thresholds may in part be determined by naturally occurring polymorphisms in genes such as *Ptch* that are involved in normal cell fate decisions.

The observation of opposing roles of *Ptch* signalling in BCC and SCC formation has implications for the development of new therapies based on targeting cancer stem cells. Although inhibitors of SHH signalling may be important therapeutic agents for BCCs[29], such inhibitors may also have additional effects in promoting cell fate decisions leading to SCCs or some other tumour types.

METHODS

See Supplementary Information for experimental details.

Mice, genotyping and linkage analysis. In brief, 440 *K5Hras* (B6FVB) F_1 backcross mice were generated for linkage analysis. *K5Hras/Ptch$^{+/-}$* or *K5Hras/K14Ptch^{FVB}* (B6FVB)F_1 mice were monitored for SCC development for at least three weeks after birth. All animal experiments were performed under protocols approved by the UCSF Laboratory Animal Resource Center.

Tail DNAs were prepared for genotyping by polymerase chain reaction with microsatellite markers. Logistic regression analysis was used to identify quantitative trait loci that control skin tumour susceptibility by the *K5Hras* transgene[25].

TaqMan analysis. Total RNAs were used to assess gene expression. Primers and probes for *Gapdh*, *Gli2* and *K14Ptch^{FVB}* are shown in Supplementary Table 2. Further details of assays and calculation of relative expression levels are shown in Supplementary Information.

Cell culture, transfections and apoptosis assays. NIH 3T3, C5N and 293T cells were grown in Dulbecco's medium supplemented with 10% fetal bovine serum and glutamine in an atmosphere of 5% CO_2. For transient transfections into NIH 3T3 cells and 293T cells, Fugene (Roche) was used in accordance with the manufacturer's instructions, and cells were collected 24 h after transfection.

C5N cells were infected with high-titre retroviral stocks produced by transient transfection of 293T amphotropic Phoenix cells. After infection with MSCV retrovirus expressing each variant of Ptch, the cells were selected in medium containing 2 μg ml^{-1} puromycin. Selected C5N cells were infected with pLXSP3 retrovirus expressing V12 HRAS and harvested 24 h after infection.

Apoptosis was detected by Annexin V-PI staining with the use of fluorescence-activated cell-sorting analysis. All experiments were repeated at least three times.

Immunoprecipitation and western blotting. Lysates (500 μg) were precleared with Protein A-Sepharose beads (Invitrogen) and incubated with 2 μg of anti-Tid1 antibody overnight. Addition of Protein A-Sepharose beads with incubation for 3–4 h followed. After the incubation, the beads were washed three times with the lysis buffer and the precipitates were subjected to western blotting. For immunoprecipitation with anti-HA antibody (Covance), a Profound HA Tag IP/Co-IP kit (Pierce) was used in accordance with the instruction manual.

For western blotting, samples were separated on Novex Nupage Tris-Glycine gels (Invitrogen) followed by electrophoretic transfer to poly(vinylidene difluoride) membranes (Millipore) and blocking in 5% nonfat milk. Immunodetection was performed by enhanced chemiluminescence detection (Amersham).

Received 25 August; accepted 27 November 2006.
Published online 17 January 2007.

1. Hennings, H. *et al.* FVB/N mice: an inbred strain sensitive to the chemical induction of squamous cell carcinomas in the skin. *Carcinogenesis* **14,** 2353–2358 (1993).
2. Hahn, H. *et al.* Mutations of the human homolog of *Drosophila* patched in the nevoid basal cell carcinoma syndrome. *Cell* **85,** 841–851 (1996).
3. Johnson, R. L. *et al.* Human homolog of patched, a candidate gene for the basal cell nevus syndrome. *Science* **272,** 1668–1671 (1996).
4. Gailani, M. R. *et al.* The role of the human homologue of *Drosophila* patched in sporadic basal cell carcinomas. *Nature Genet.* **14,** 78–81 (1996).
5. Dahmane, N., Lee, J., Robins, P., Heller, P. & Ruiz i Altaba, A. Activation of the transcription factor Gli1 and the Sonic hedgehog signalling pathway in skin tumours. *Nature* **389,** 876–881 (1997).
6. Brown, K., Strathdee, D., Bryson, S., Lambie, W. & Balmain, A. The malignant capacity of skin tumours induced by expression of a mutant H-ras transgene depends on the cell type targeted. *Curr. Biol.* **8,** 516–524 (1998).
7. Li, X. *et al. Ex vivo* culture of Fancc$^{-/-}$ stem/progenitor cells predisposes cells to undergo apoptosis, and surviving stem/progenitor cells display cytogenetic abnormalities and an increased risk of malignancy. *Blood* **105,** 3465–3471 (2005).
8. Aszterbaum, M. *et al.* Ultraviolet and ionizing radiation enhance the growth of BCCs and trichoblastomas in patched heterozygous knockout mice. *Nature Med.* **5,** 1285–1291 (1999).
9. Unden, A. B. *et al.* Mutations in the human homologue of *Drosophila* patched (PTCH) in basal cell carcinomas and the Gorlin syndrome: different *in vivo* mechanisms of PTCH inactivation. *Cancer Res.* **56,** 4562–4565 (1996).
10. Arbeit, J. M., Munger, K., Howley, P. M. & Hanahan, D. Progressive squamous epithelial neoplasia in K14-human papillomavirus type 16 transgenic mice. *J. Virol.* **68,** 4358–4368 (1994).
11. Chin, L. *et al.* Essential role for oncogenic Ras in tumour maintenance. *Nature* **400,** 468–472 (1999).
12. Weinstein, I. B. Cancer. Addiction to oncogenes—the Achilles heal of cancer. *Science* **297,** 63–64 (2002).
13. Thibert, C. *et al.* Inhibition of neuroepithelial patched-induced apoptosis by sonic hedgehog. *Science* **301,** 843–846 (2003).
14. Barnes, E. A., Kong, M., Ollendorff, V. & Donoghue, D. J. Patched1 interacts with cyclin B1 to regulate cell cycle progression. *EMBO J.* **20,** 2214–2223 (2001).
15. Goodrich, L. V., Milenković, L., Higgins, K. M. & Scott, M. P. Altered neural cell fates and medulloblastoma in mouse patched mutants. *Science* **277,** 1109–1113 (1997).
16. Guerrero, I. & Ruiz i Altaba, A. Development. Longing for ligand: hedgehog, patched, and cell death. *Science* **301,** 774–776 (2003).
17. Hooper, J. E. & Scott, M. P. Communicating with Hedgehogs. *Nature Rev. Mol. Cell Biol.* **6,** 306–317 (2005).
18. Syken, J., De-Medina, T. & Münger, K. *TID1*, a human homolog of the *Drosophila* tumor suppressor *l(2)tid*, encodes two mitochondrial modulators of apoptosis with opposing functions. *Proc. Natl Acad. Sci. USA* **96,** 8499–8504 (1996).
19. Canamasas, I., Debes, A., Natali, P. G. & Kurzik-Dumke, U. Understanding human cancer using *Drosophila*: Tid47, a cytosolic product of the *DnaJ*-like tumor suppressor gene *l(2)Tid*, is a novel molecular partner of Patched related to skin cancer. *J. Biol. Chem.* **278,** 30952–30960 (2003).
20. Trentin, G. A. *et al.* A mouse homologue of the *Drosophila* tumor suppressor *l(2)tid* gene defines a novel Ras GTPase-activating protein (RasGAP)-binding protein. *J. Biol. Chem.* **276,** 13087–13095 (2001).
21. Sarkar, S. *et al.* hTid-1, a human DnaJ protein, modulates the interferon signaling pathway. *J. Biol. Chem.* **276,** 49034–49042 (2001).
22. Tarunina, M. *et al.* Functional genetic screen for genes involved in senescence: role of Tid1, a homologue of the *Drosophila* tumor suppressor l(2)tid, in senescence and cell survival. *Mol. Cell. Biol.* **24,** 10792–10801 (2004).
23. Cheng, H. *et al.* Molecular mechanism of hTid-1, the human homolog of *Drosophila* tumor suppressor l(2)Tid, in the regulation of NF-κB activity and suppression of tumor growth. *Mol. Cell. Biol.* **25,** 44–59 (2005).
24. Mao, J. H. & Balmain, A. Genomic approaches to identification of tumour susceptibility genes using mouse models. *Curr. Opin. Genet. Dev.* **13,** 14–19 (2003).
25. Nagase, H. *et al.* Distinct genetic loci control development of benign and malignant skin tumours in mice. *Nature Genet.* **10,** 424–429 (1995).
26. Perez-Losada, J. & Balmain, A. Stem-cell hierarchy in skin cancer. *Nature Rev. Cancer* **3,** 434–443 (2003).
27. Quintanilla, M., Brown, K., Ramsden, M. & Balmain, A. Carcinogen-specific mutation and amplification of Ha-ras during mouse skin carcinogenesis. *Nature* **322,** 78–80 (1986).
28. Janes, K. A. *et al.* A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis. *Science* **310,** 1646–1653 (2005).
29. Taipale, J. *et al.* Effects of oncogenic mutations in Smoothened and Patched can be reversed by cyclopamine. *Nature* **406,** 1005–1009 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Curran for providing *Ptch*$^{+/-}$ mice, E. Bailey for providing *Ptch* constructs, U. Kurzik-Dumke and M. Rozakis-Adcock for anti-Tid1 antibodies, E. Epstein and numerous colleagues for discussions and comments, R. del Rosario for assistance with animal husbandry, and N. Killeen and the UCSF transgenic core facility for assistance in the generation of transgenic mice. This work was supported by grants from the National Cancer Institute and the Department of Energy Low Dose Program. A.B. acknowledges support from the Barbara Bass Bakar Chair of Cancer Genetics.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.B. (abalmain@cc.ucsf.edu).

Vol 445|15 February 2007|doi:10.1038/nature05479

LETTERS

Regulatory T-cell functions are subverted and converted owing to attenuated Foxp3 expression

Yisong Y. Wan[1] & Richard A. Flavell[1,2]

The naturally occurring regulatory T cell (T_R) is the pivotal cell type that maintains self-tolerance and exerts active immune suppression. The development and function of T_R cells is controlled by Foxp3 (refs 1, 2), a lack of which results in loss of T_R cells and massive multi-organ autoimmunity in scurfy mice and *IPEX* (immune dysregulation, polyendocrinopathy, enteropathy, X-linked) patients[3,4]. It is generally thought that, through a binary mechanism, Foxp3 expression serves as an on-and-off switch to regulate positively the physiology of T_R cells; however, emerging evidence associates decreased Foxp3 expression in T_R cells with various immune disorders[5–7]. We hypothesized that Foxp3 regulates T_R cell development and function in a dose-dependent, non-binary manner, and that decreased Foxp3 expression can cause immune disease. Here, by generating a mouse model in which endogenous *Foxp3* gene expression is attenuated in T_R cells, we show that decreased Foxp3 expression results in the development of an aggressive autoimmune syndrome similar to that of scurfy mice, but does not affect thymic development, homeostatic expansion/maintenance or transforming-growth-factor-β-induced *de novo* generation of Foxp3-expressing cells. The immune-suppressive activities of T cells with attenuated Foxp3 expression were nearly abolished *in vitro* and *in vivo*, whereas their anergic properties *in vitro* were maintained. This was accompanied by decreased expression of T_R cell 'signature genes'. Notably, T cells expressing decreased Foxp3 preferentially became T-helper 2 (T_H2)-type effectors even in a T_H1-polarizing environment. These cells instructed T_H2 differentiation of conventional T cells, which contributed to the immune diseases observed in these mice. Thus, decreased Foxp3 expression causes immune disease by subverting the suppressive function of T_R cells and converting T_R cells into effector cells; these findings are important for understanding the regulation of T_R cell function and the aetiology of various human immune diseases.

T_R cells, a central component for immune suppression, are critical for establishing self-tolerance, controlling inflammatory responses and maintaining immune homeostasis[8,9]. Foxp3, an X-chromosome-linked factor that controls T_R cell development and function[1,2], is generally thought to control positively the functions of T_R cells in a binary fashion, as Foxp3 expression is sufficient to specify immune-suppressive activities in conventional T cells[1,2,10]. Thus, current efforts are focused on associating abnormal numbers of T_R cells with immune disorders. However, the quality of T_R cells is also critical for their function[11]. We observed lowered levels of Foxp3 in intra-islet T_R cells compared with T_R cells from other peripheral lymphoid organs in diabetic NOD mice (Fig. 1a), whereas the frequencies of Foxp3-expressing T_R cells among different compartments were comparable (data not shown). However, such a specific decrease in Foxp3 expression was not observed in non-diabetes-prone C57BL/6 mice (Supplementary Fig. 1). We hypothesized that one of the quality control mechanisms for T_R cells is through tuning the expression levels of Foxp3, and that decreased Foxp3 expression can cause immune disease. To test this, we generated a mouse model, where attenuated expression of the endogenous *Foxp3* gene was achieved by a targeted gene 'knock-in' approach, allowing us to investigate the effects of decreased Foxp3 expression on T_R cell function and to provide potential mechanistic explanations for the aetiologies of certain human immune disorders. In this model, a gene cassette co-expressing luciferase and enhanced green fluorescent protein (eGFP)—the translation of which was under the control of two tandem internal ribosomal entry sites (IRES)—was inserted into the 3′-untranslated region (UTR) of the endogenous *Foxp3* locus of C57BL/6 mice to generate a Foxp3-IRES-luciferase-IRES-eGFP (*FILIG*) allele (Supplementary Fig. 2). Using a similar approach, we have previously generated a knock-in mouse model where Foxp3-expressing cells are marked by the co-expression of a monomeric red fluorescent protein (RFP)[12]. Such Foxp3-IRES-mRFP (*FIR*) mice can be used to isolate wild-type T_R cells based on RFP expression. In the following experiments, hemizygous male *FIR* mice (*FIR/Y*), heterozygous female *FIR* mice (*FIR/+*) or homozygous female *FIR* mice (*FIR/FIR*) and T cells from these mice were used as wild-type controls where appropriate.

FILIG mice were born at a mendelian ratio. Heterozygous *FILIG* female mice (*FILIG/+*) were fertile and phenotypically normal; however, hemizygous *FILIG* male mice (*FILIG/Y*) were barren and runted (Fig. 1b). Over 50% of *FILIG/Y* mice developed scaly skin (data not shown) and nearly all of them developed eyelid defects resembling blepharitis, a T_H2 disorder, at around 4 weeks of age (Supplementary Fig. 3). By 3 months of age, all the *FILIG/Y* mice succumbed to an aggressive lymphoproliferative autoimmune syndrome, manifested by enlarged spleens and lymph nodes (Fig. 1b), infiltration of lymphocytes into non-lymphoid organs, drastically increased serum levels of auto-antibodies (Supplementary Fig. 4), and activated $CD4^+$ and $CD8^+$ T cells (Supplementary Fig. 5). Overall, *FILIG/Y* mice displayed phenotypes reminiscent of scurfy mice[13] and T-cell-specific *Foxp3* knockout mice[1]. To investigate whether transcription of the endogenous *Foxp3* gene was abolished in *FILIG* mice, we first detected luciferase expression by live imagining. In *FILIG/+* mice, cells expressing luciferase were concentrated in lymphoid organs (Fig. 1c). However, in *FILIG/Y* mice, these cells were detected in lymphoid as well as non-lymphoid organs (Fig. 1c), suggesting that the endogenous *Foxp3* gene was transcribed in *FILIG* lymphocytes, and that *FILIG* lymphocytes infiltrated non-lymphoid organs in *FILIG/Y* mice. By flow cytometry, we detected GFP expression only in $CD4^+$ T cells (Supplementary Fig. 6). Notably, among $CD4^+$ T cells, there was a higher percentage of GFP^+ cells in *FILIG/Y* mice compared with RFP^+ cells in *FIR/Y* mice, whereas a lower percentage of GFP^+ cells was detected in *FILIG/+* mice compared with RFP^+ cells in *FIR/+* mice (Fig. 1d). By intracellular staining, we detected

[1]Section of Immunobiology, Yale University School of Medicine, 300 Cedar Street, and [2]Howard Hughes Medical Institute, New Haven, Connecticut 06520, USA.

Foxp3 expression in GFP$^+$ CD4$^+$ T cells from *FILIG/Y* and *FILIG/+* mice, but not in GFP$^-$ CD4$^+$ T cells from *FILIG/Y* mice (Fig. 1e). Therefore, GFP expression reflected Foxp3 expression with high fidelity in *FILIG* mice. Compared with wild-type T_R cells, GFP$^+$ cells from *FILIG* mice expressed 5–10-fold less Foxp3. Thus, although Foxp3 was expressed at decreased levels, *FILIG* mice generated Foxp3-expressing CD4$^+$ T cells that were faithfully marked by GFP expression. Compared with wild-type T_R cells, the surface expression and messenger RNA levels of signature genes for T_R cells (such as *CD25*, *Ctla4* and *Gitr*[14–16]) were decreased in GFP$^+$ CD4$^+$ T cells from *FILIG/+* mice (Fig. 1f, g). The exact mechanism by which attenuated Foxp3 expression was achieved in *FILIG* CD4$^+$ T cells is unclear. However, four AU-rich elements (ARE), the presence of which in the 3′-UTR of a gene is known to destabilize mRNA[17], were found dispersed in the luciferase complementary DNA. Therefore, reduced Foxp3 expression in *FILIG* CD4$^+$ T cells is probably due to mRNA destabilization caused by localization of the luciferase sequence in the 3′-UTR of the *Foxp3* mRNA. As a consequence of random X-chromosome inactivation, the *FILIG/+* female mice contain a mixture of Foxp3$^+$ cells: those with wild-type levels of Foxp3 expression and those having low levels due to the gene knock-in approach; this mixture of cells probably accounts for lack of overt disease in these mice.

The *FILIG* model allowed us to investigate further which biological functions of T_R cells are altered owing to decreased Foxp3 expression. To assess whether attenuated Foxp3 expression affected T_R cell development, the percentages of Foxp3-expressing cells (Foxp3$^+$) among CD4$^+$ single positive thymocytes from 10- and 16-day-old *FILIG/Y* and *FIR/Y* mice were determined and compared, but no difference was observed (Fig. 2a). Whether homeostatic expansion/maintenance of Foxp3$^+$ cells was affected by attenuated Foxp3 expression was addressed by transferring cell mixtures combining conventional CD4$^+$ T cells (RFP$^-$) from *FIR/FIR* mice with Foxp3$^+$ (RFP$^+$ or

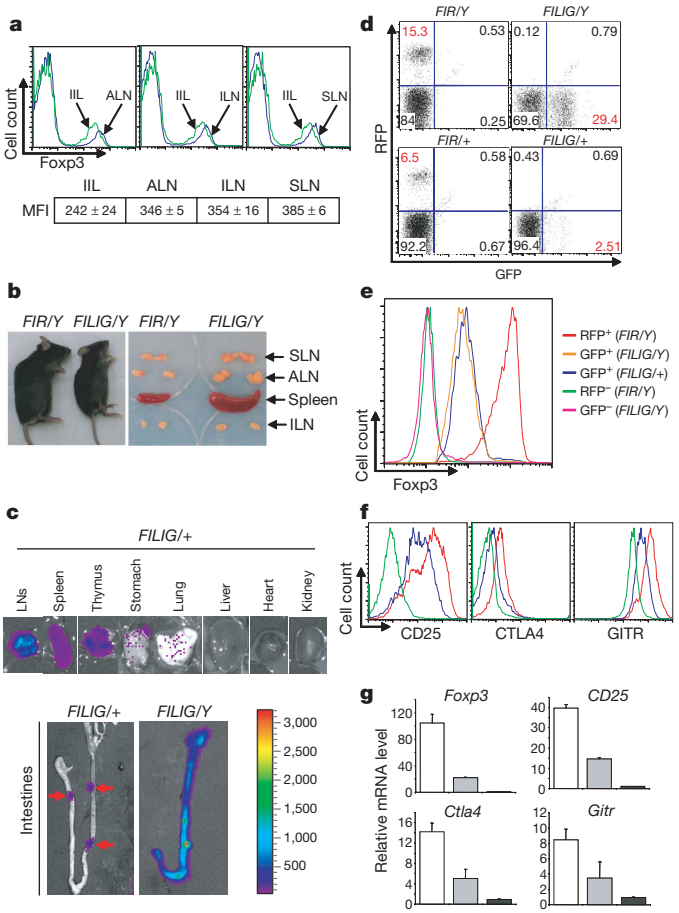


Figure 1 | Attenuated Foxp3 expression in T_R cells results in an aggressive autoimmune syndrome and altered surface properties of T_R cells. **a**, Flow cytometry for Foxp3 expression in CD4$^+$ T cells isolated from intra-islet lymphocytes (IIL; green line), and axillary (ALN), superficial-inguinal (ILN) and submandibular (SLN) lymph nodes (blue lines) from diabetic NOD mice. Mean fluorescence intensity (MFI) of Foxp3 staining among Foxp3$^+$ cells is indicated. Data are mean ± s.d. of two samples with one sample combining lymphocytes from three mice. **b**, Size comparison of 2-month-old *FIR/Y* and *FILIG/Y* mice (left panel); sizes of submandibular (SLN), axillary (ALN) and superficial-inguinal (ILN) lymph nodes and spleens are also shown (right panel). **c**, Localization of luciferase-expressing cells in various organs in *FILIG/+* and *FILIG/Y* mice. The colour scale with corresponding count units is shown. **d**, The percentage (highlighted in red) of GFP$^+$ cells among CD4$^+$ T cells in *FILIG/Y* and *FILIG/+* mice and of RFP$^+$ cells in *FIR/Y* and *FIR/+* control mice is shown. Results representative of at least ten experiments are shown. **e**, Flow cytometry of Foxp3 in sorted GFP$^+$ CD4$^+$ T cells from *FILIG/+* mice; GFP$^+$ and GFP$^-$ CD4$^+$ T cells from *FILIG/Y* mice; and RFP$^+$ and RFP$^-$ CD4$^+$ T cells from *FIR/Y* mice. Results representative of four experiments are shown. **f**, Surface expression of CD25, CTLA4 and GITR on GFP$^+$ CD4$^+$ T cells from *FILIG/+* mice (blue lines); RFP$^+$ CD4$^+$ T cells from *FIR/+* mice (red lines); and RFP$^-$ CD4$^+$ T cells from *FIR/FIR* mice (green lines). Results are representative of three experiments. **g**, Relative mRNA levels of *Foxp3*, *CD25*, *Ctla4* and *Gitr* in sorted GFP$^+$ CD4$^+$ T cells (grey bars) from *FILIG/+* mice, and RFP$^+$ (white bars) and RFP$^-$ (black bars) CD4$^+$ T cells from *FIR/FIR* mice. Data are mean ± s.d. of combined results from three experiments. The differences observed among samples for different genes are statistically significant as determined by the 'Statistical analysis' section of the Methods.

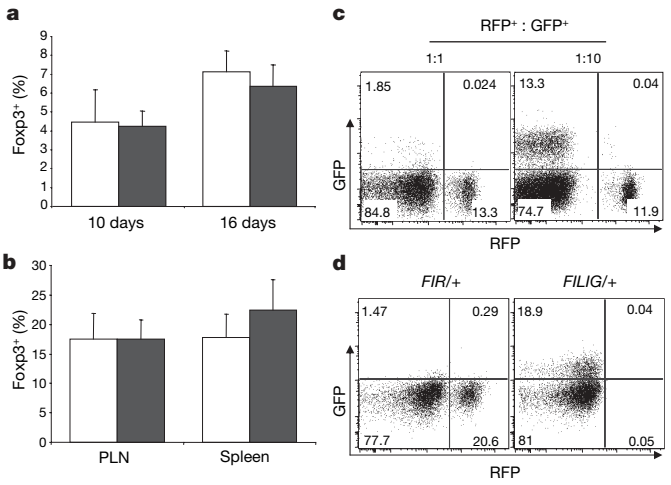


Figure 2 | Thymic development, homeostatic expansion/maintenance and TGF-β-induced *de novo* generation of T_R cells are normal when Foxp3 expression is decreased. **a**, The percentage of Foxp3$^+$ CD4$^+$ single positive thymocytes in 10- and 16-day-old *FIR/Y* (white bars) and *FILIG/Y* mice (black bars) is shown. Data are mean ± s.d. of results from four mice for each time point. **b**, Sorted RFP$^-$ CD4$^+$ T cells from *FIR/FIR* mice were mixed with either RFP$^+$ CD4$^+$ T cells from *FIR/+* mice (white bars) or GFP$^+$ CD4$^+$ T cells from *FILIG/+* mice (black bars) at a ratio of 2:1 and then transferred into female *Rag1*$^{-/-}$ mice. The percentage of Foxp3$^+$ (RFP$^+$ or GFP$^+$) CD4$^+$ T cells from different origins in different lymphoid organs was determined by flow cytometry 9 weeks later. PLN, peripheral lymph nodes and spleen. Data are mean ± s.d. of results from six mice of one experiment representative of two. **c**, RFP$^+$ and GFP$^+$ CD4$^+$ T cells were sorted from *FIR/+* and *FILIG/+* mice respectively, and then combined at the indicated ratios. 1×10^5 cell mixtures and 2×10^5 RFP$^-$ CD4$^+$ T cells sorted from *FIR/FIR* mice were co-transferred into *Rag1*$^{-/-}$ mice. The distribution of Foxp3$^+$ *FIR* (RFP$^+$) and Foxp3$^+$ *FILIG* (GFP$^+$) cells in the recipient mice was determined by flow cytometry 9 weeks later. Results representative of six mice are shown. **d**, CD4$^+$CD25$^-$ T cells that were negative for RFP and GFP were sorted from *FIR/+* and *FILIG/+* mice respectively, and then activated in the presence of TGF-β1. Foxp3 expression was measured by the expression of RFP and GFP. Results representative of three experiments are shown.

GFP$^+$) CD4$^+$ T cells from either *FIR*/+ or *FILIG*/+ mice into *Rag1*$^{-/-}$ mice. RFP$^+$ and GFP$^+$ CD4$^+$ T cells in the recipient mice were detected 9 weeks after transfer (Supplementary Fig. 7); the percentage of RFP$^+$ and GFP$^+$ cells among transferred CD4$^+$ T cells was comparable (Fig. 2b), suggesting that decreased Foxp3 expression did not result in intrinsic defects of the homeostatic expansion/maintenance of GFP$^+$ *FILIG* cells. However, when RFP$^+$ and GFP$^+$ CD4$^+$ T cells were co-transferred into the same hosts in the presence of conventional CD4$^+$ T cells, GFP$^+$ cells competed poorly with RFP$^+$ cells (Fig. 2c), in agreement with the observation that a lower than expected percentage of GFP$^+$ cells was found in the *FILIG*/+ mice.

The fact that CD25 expression on GFP$^+$ cells from *FILIG*/+ mice was decreased relative to that on RFP$^+$ cells from *FIR*/+ mice could account for this phenomenon, as T_R cell maintenance is dependent on interleukin-2 (IL-2) signalling[18–20]. Extra-thymic generation of Foxp3$^+$ T cells can be promoted *in vitro* by transforming growth factor-β (TGF-β)[12,21,22]. TGF-β induced *de novo* Foxp3 expression in *FILIG* CD4$^+$ T cells to a similar extent as in *FIR* CD4$^+$ T cells (Fig. 2d).

In vitro, anergy and immune-suppressive activities are two defining properties for T_R cells that are thought to go hand-in-hand[23]. Notably, upon T-cell-receptor (TCR) stimulation *in vitro*, although GFP$^+$ cells from *FILIG* mice remained anergic (Fig. 3a and

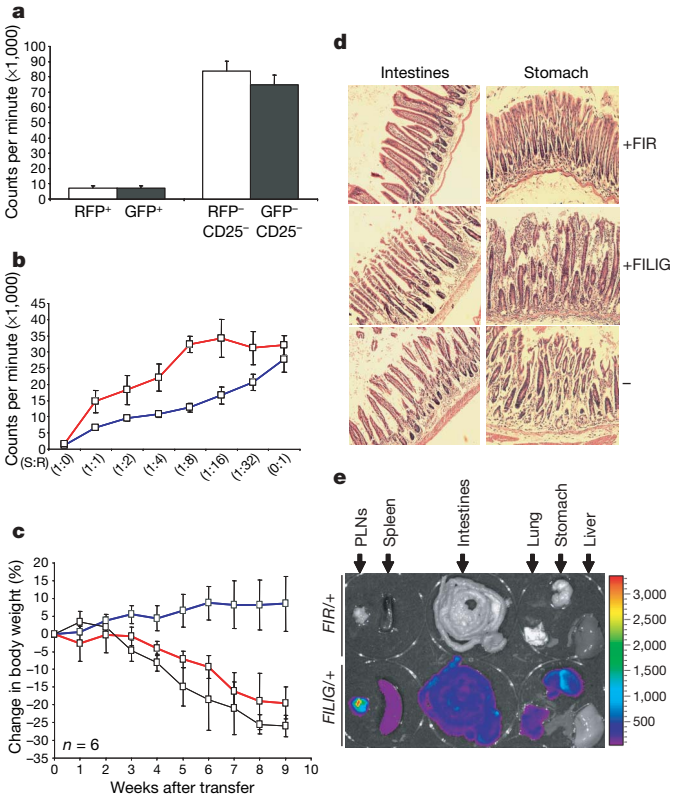


Figure 3 | Attenuation of Foxp3 expression abrogated the immune-suppressive but not hypoproliferative activities of T_R cells. **a**, RFP$^+$, GFP$^+$, RFP$^-$CD25$^-$ and GFP$^-$CD25$^-$ CD4$^+$ T cells were sorted from *FIR*/+ (white bars) and *FILIG*/+ (black bars) mice. Purified cells were activated *in vitro*, and cell proliferation was determined by a T-cell proliferation assay. Data are mean ± s.d. of triplicates done in one experiment representative of three. **b**, Suppression assay performed using sorted RFP$^+$ CD4$^+$ T cells from *FIR*/+ mice (blue line) or GFP$^+$ CD4$^+$ T cells from *FILIG*/+ mice (red line) as suppressor cells (S) and RFP$^-$ CD4$^+$ T cells sorted from *FIR/FIR* mice as responder cells (R). Data are mean ± s.d. of triplicates done in one experiment representative of three. **c**, *Rag1*$^{-/-}$ mice were transferred with RFP$^-$ CD4$^+$ T cells sorted from *FIR/FIR* mice alone (black line), or with T-cell mixtures containing one-third of RFP$^+$ CD4$^+$ T cells sorted from *FIR*/+ mice (blue line) or GFP$^+$ CD4$^+$ T cells sorted from *FILIG*/+ mice (red line) and two-thirds of RFP$^-$ CD4$^+$ T cells sorted from *FIR/FIR* mice. The percentage of body weight change in the recipient mice was determined weekly for 9 weeks after transfer. Data are mean ± s.d. of six mice from one experiment representative of two. **d**, Intestines and stomach were removed from recipient mice at the end of the experiments described in **c**. Haematoxylin-and-eosin staining was performed to detect pathological changes in mice that received conventional CD4$^+$ T cells together with RFP$^+$ CD4$^+$ T cells from *FIR*/+ (+FIR) or GFP$^+$ CD4$^+$ T cells from *FILIG*/+ (+FILIG) mice; or without Foxp3$^+$ cells (−). **e**, At the end of experiments described in **c**, localization of transferred GFP$^+$ *FILIG* cells in different organs (as indicated) was determined by live imaging. The colour scale with corresponding count units is shown. As a control, organs from a mouse that received RFP$^+$ *FIR* cells are shown.

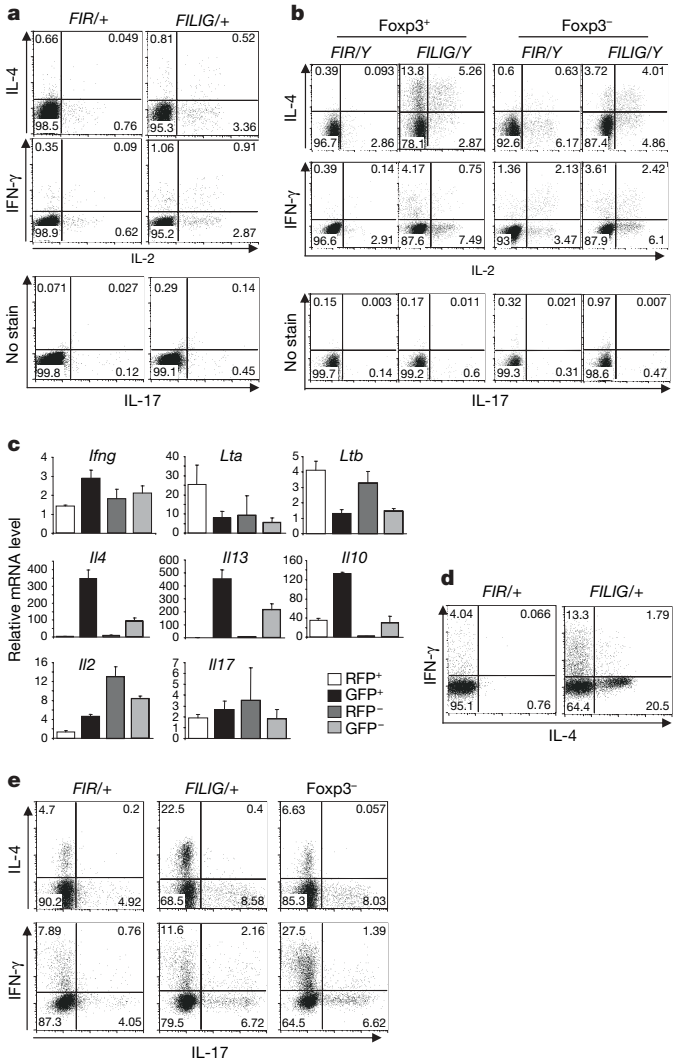


Figure 4 | T_R cells converted into T_H2 effector cells owing to decreased Foxp3 expression. **a**, **b**, The production of IL-2, IL-4, IL-17 and IFN-γ in RFP$^+$ CD4$^+$ T cells sorted from *FIR*/+ mice and in GFP$^+$ CD4$^+$ T cells sorted from *FILIG*/+ mice (**a**), or in RFP$^+$ and RFP$^-$ CD4$^+$ T cells from *FIR/Y* mice and GFP$^+$ and GFP$^-$ CD4$^+$ T cells from *FILIG/Y* mice (**b**) is shown. **c**, Relative mRNA levels of *Ifng*, *Lta* (lymphotoxin A), *Ltb* (lymphotoxin B), *Il4*, *Il13*, *Il10*, *Il2* and *Il17* among RFP$^+$, RFP$^-$, GFP$^+$ and GFP$^-$ CD4$^+$ T cells isolated from male *FIR/Y* and *FILIG/Y* mice are shown. Data are mean ± s.d. from two experiments. **d**, At the end of the experiments described in Fig. 3c, The expression of IL-4 and IFN-γ was examined in RFP$^+$ *FIR*/+ and GFP$^+$ *FILIG*/+ cells that were recovered from recipient mice. **e**, At the end of the experiments described in Fig. 2c (right panel), the expression of IL-4, IL-17 and IFN-γ was determined in co-existing RFP$^+$ *FIR*/+ cells and GFP$^+$ *FILIG*/+ cells as well as in RFP$^-$GFP$^-$ *FIR/FIR* cells. All results are representative of at least three experiments unless stated otherwise.

Supplementary Fig. 8), their immune-suppressive activities were greatly impaired (Fig. 3b). Thus, anergy and immune suppression are two separable properties of T_R cells that are affected differentially by Foxp3 expression level. The immune-suppressive activities of Foxp3$^+$ CD4$^+$ T cells from *FILIG* mice were also abolished *in vivo*, because, unlike wild-type T_R cells, co-transferred GFP$^+$ cells from *FILIG/+* mice did not prevent conventional CD4$^+$ T-cell-elicited weight loss (Fig. 3c) or the immune pathologies in the intestines and stomach (Fig. 3d) of *Rag1*$^{-/-}$ recipient mice. This was not due to defective migration of transferred cells. In fact, adoptively transferred GFP$^+$ *FILIG* cells are found in lymphoid as well as non-lymphoid organs (Fig. 3e).

Loss of the suppressive activity of T_R cells could be sufficient to cause an aggressive autoimmune syndrome in *FILIG/Y* mice and the wasting disease observed in the aforementioned transfer model. Nevertheless, we further investigated whether Foxp3$^+$ *FILIG* cells developed effector functions that could contribute to any of these immune disorders. On the basis of cytokine production profiles, three types of effector T cells, T_H1, T_H2 and T_H17, have been described[24,25]. Although T_R cells bear self-reactive TCRs[26,27], GFP$^+$ cells from healthy *FILIG/+* mice did not exhibit an effector cell phenotype (Fig. 4a), suggesting that they did not spontaneously activate and convert into effector cells when substantial numbers of wild-type T_R cells were present. Notably, however, a large portion of GFP$^+$ cells from diseased *FILIG/Y* mice produced IL-4, whereas the percentage of cells expressing IL-2, interferon-γ (IFN-γ) or IL-17 was only modestly increased compared to wild-type cells (Fig. 4b). The percentage of IL-4-producing cells also preferentially increased in Foxp3$^-$ (GFP$^-$) CD4$^+$ T cells from *FILIG/Y* mice, consistent with the T_H2 disorder observed in these mice. Cytokine mRNA levels were also determined and agreed with aforementioned results (Fig. 4c). In addition, compared with wild-type T_R cells, the percentage of IL-4-producing GFP$^+$ *FILIG* cells showed an approximately 30-fold increase upon adoptive transfer, whereas that of IFN-γ-producing GFP$^+$ *FILIG* cells showed only an approximately 3-fold increase (Fig. 4d). In some experiments, we also noticed that the percentage of IL-4- but not IFN-γ-producing cells was substantially increased even among transferred wild-type Foxp3$^+$ *FIR* cells that expressed lower levels of Foxp3 (Supplementary Fig. 9), suggesting that IL-4 production can be induced in wild-type T_R cells expressing low levels of Foxp3 *in vivo*. To compare Foxp3$^+$ *FIR* and *FILIG* CD4$^+$ T cells directly in the same physiological environment, CD4$^+$ T cells that were RFP$^+$ from *FIR/+* mice, GFP$^+$ from healthy *FILIG/+* mice, and RFP$^-$ from *FIR/FIR* mice, were mixed at a ratio of 1:10:20 and then transferred into *Rag1*$^{-/-}$ hosts. At the time of transfer, none of these cells was producing substantial amounts of effector cytokines (Fig. 4a and data not shown). Recipient mice developed wasting disease 9 weeks after transfer (data not shown), possibly owing to the fact that insufficient numbers of wild-type T_R cells were transferred. Compared with coexisting RFP$^+$ wild-type T_R cells, the fraction of IL-4-producing cells among GFP$^+$ *FILIG* cells increased greatly, whereas that of IFN-γ- or IL-17-producing cells did not change substantially, although coexisting Foxp3$^-$ cells produced large quantities of IFN-γ, thereby providing a T_H1-polarizing condition (Fig. 4e). Moreover, GFP$^+$ *FILIG* cells potently induced T_H2 differentiation of conventional CD4$^+$ T cells *in vitro* and *in vivo*, potentially through IL-4. Intriguingly, IL-4 production by GFP$^+$ *FILIG* cells was not affected by the T_H1-polarizing environments imposed by co-cultured cells (Supplementary Fig. 10).

Decreased Foxp3 expression is associated with human immune disorders[5–7]. Our genetic evidence convincingly shows that decreased Foxp3 expression can cause defective suppressive function of T_R cells and their conversion into effector cells, which contribute to rather than inhibit immune diseases. This provides an important mechanistic explanation for the aetiology of immunopathology in our mouse model and potentially various human immune diseases.

METHODS

Mice and adoptive transfer assays. *FIR*, *FILIG*, *NOD*, C57BL/6 and *Rag1*$^{-/-}$ (C57BL/6 background) mice were kept under specific pathogen-free conditions in the animal care facility at Yale University. All mouse experiments were approved by the Institutional Animal Care and Use Committee of Yale University. For adoptive transfer assays, conventional (RFP$^-$) CD4$^+$ T cells from *FIR/FIR* mice, RFP$^+$ CD4$^+$ T cells from *FIR/+* mice or GFP$^+$ CD4$^+$ T cells from *FILIG/+* mice were sorted by fluorescence-activated cell sorting (FACS). Sorted cells were either transferred alone or mixed at different ratios as elaborated in the text or figure legends. A total of 3×10^5 cells were transferred into female *Rag1*$^{-/-}$ mice via retro-orbital injection. Mice were weighed every week thereafter and killed 9–10 weeks after transfer.

Generation of *FILIG* mice. *FILIG* knock-in mice were generated according to the protocols described for generating *FIR* mice[12]. A gene cassette encoding IRES-luciferase-IRES-eGFP instead of IRES-RFP was inserted into an *Ssp1* site. The floxed neomycin cassette was deleted *in vitro* by transfecting Cre-expressing plasmid into Bruce-4 embryonic stem cells originating from C57BL/6 mice.

Live imaging and histology. For live imaging analysis, mice were injected intraperitoneally with 3 mg luciferin (Xenogen) per mouse and then killed. Different lymphoid and non-lymphoid organs were surgically removed. Luciferase-expressing organs were visualized by IVIS Imaging System (Xenogen) as per the manufacturer's protocols. For histology analysis, organs from mice were removed and fixed in S.T.F. fixatives (Streck) for 24 h. Preparation of the slides, sectioning and haematoxylin-and-eosin staining were performed by the Yale histology and pathology laboratory.

Statistical analysis. Data from at least three sets of samples were used for statistical analysis. Mean ± s.d. are shown. Statistical significance was calculated by Student's *t*-test. A *P*-value of less than 0.05 was considered significant.

Received 31 October; accepted 27 November 2006.
Published online 14 January 2007.

1. Fontenot, J. D., Gavin, M. A. & Rudensky, A. Y. Foxp3 programs the development and function of CD4$^+$CD25$^+$ regulatory T cells. *Nature Immunol.* **4,** 330–336 (2003).
2. Hori, S., Nomura, T. & Sakaguchi, S. Control of regulatory T cell development by the transcription factor Foxp3. *Science* **299,** 1057–1061 (2003).
3. Wildin, R. S. *et al.* X-linked neonatal diabetes mellitus, enteropathy and endocrinopathy syndrome is the human equivalent of mouse scurfy. *Nature Genet.* **27,** 18–20 (2001).
4. Brunkow, M. E. *et al.* Disruption of a new forkhead/winged-helix protein, scurfin, results in the fatal lymphoproliferative disorder of the scurfy mouse. *Nature Genet.* **27,** 68–73 (2001).
5. Miura, Y. *et al.* Association of Foxp3 regulatory gene expression with graft-versus-host disease. *Blood* **104,** 2187–2193 (2004).
6. Balandina, A., Lecart, S., Dartevelle, P., Saoudi, A. & Berrih-Aknin, S. Functional defect of regulatory CD4$^+$CD25$^+$ T cells in the thymus of patients with autoimmune myasthenia gravis. *Blood* **105,** 735–741 (2005).
7. Huan, J. *et al.* Decreased FOXP3 levels in multiple sclerosis patients. *J. Neurosci. Res.* **81,** 45–52 (2005).
8. Sakaguchi, S. Regulatory T cells: key controllers of immunologic self-tolerance. *Cell* **101,** 455–458 (2000).
9. Shevach, E. M. Regulatory T cells in autoimmunity. *Annu. Rev. Immunol.* **18,** 423–449 (2000).
10. Khattri, R., Cox, T., Yasayko, S. A. & Ramsdell, F. An essential role for Scurfin in CD4$^+$CD25$^+$ T regulatory cells. *Nature Immunol.* **4,** 337–342 (2003).
11. Fontenot, J. D. & Rudensky, A. Y. A well adapted regulatory contrivance: regulatory T cell development and the forkhead family transcription factor Foxp3. *Nature Immunol.* **6,** 331–337 (2005).
12. Wan, Y. Y. & Flavell, R. A. Identifying Foxp3-expressing suppressor T cells with a bicistronic reporter. *Proc. Natl Acad. Sci. USA* **102,** 5126–5131 (2005).
13. Lyon, M. F., Peters, J., Glenister, P. H., Ball, S. & Wright, E. The scurfy mouse mutant has previously unrecognized hematological abnormalities and resembles Wiskott-Aldrich syndrome. *Proc. Natl Acad. Sci. USA* **87,** 2433–2437 (1990).
14. Sakaguchi, S., Sakaguchi, N., Asano, M., Itoh, M. & Toda, M. Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor α-chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. *J. Immunol.* **155,** 1151–1164 (1995).
15. Takahashi, T. *et al.* Immunologic self-tolerance maintained by CD25$^+$CD4$^+$ regulatory T cells constitutively expressing cytotoxic T lymphocyte-associated antigen 4. *J. Exp. Med.* **192,** 303–310 (2000).
16. Ono, M., Shimizu, J., Miyachi, Y. & Sakaguchi, S. Control of autoimmune myocarditis and multiorgan inflammation by glucocorticoid-induced TNF receptor family-related protein^{high}, Foxp3-expressing CD25$^+$ and CD25$^-$ regulatory T cells. *J. Immunol.* **176,** 4748–4756 (2006).
17. Chen, C. Y. & Shyu, A. B. AU-rich elements: characterization and importance in mRNA degradation. *Trends Biochem. Sci.* **20,** 465–470 (1995).
18. Furtado, G. C., Curotto de Lafaille, M. A., Kutchukhidze, N. & Lafaille, J. J. Interleukin 2 signaling is required for CD4$^+$ regulatory T cell function. *J. Exp. Med.* **196,** 851–857 (2002).

19. Fontenot, J. D., Rasmussen, J. P., Gavin, M. A. & Rudensky, A. Y. A function for interleukin 2 in Foxp3-expressing regulatory T cells. *Nature Immunol.* **6,** 1142–1151 (2005).
20. D'Cruz, L. M. & Klein, L. Development and function of agonist-induced $CD25^+Foxp3^+$ regulatory T cells in the absence of interleukin 2 signaling. *Nature Immunol.* **6,** 1152–1159 (2005).
21. Chen, W. *et al.* Conversion of peripheral $CD4^+CD25^-$ naive T cells to $CD4^+CD25^+$ regulatory T cells by TGF-β induction of transcription factor Foxp3. *J. Exp. Med.* **198,** 1875–1886 (2003).
22. Bettelli, E. *et al.* Reciprocal developmental pathways for the generation of pathogenic effector T_H17 and regulatory T cells. *Nature* **441,** 235–238 (2006).
23. Sakaguchi, S. Naturally arising $CD4^+$ regulatory T cells for immunologic self-tolerance and negative control of immune responses. *Annu. Rev. Immunol.* **22,** 531–562 (2004).
24. Mosmann, T. R. & Coffman, R. L. TH1 and TH2 cells: different patterns of lymphokine secretion lead to different functional properties. *Annu. Rev. Immunol.* **7,** 145–173 (1989).
25. Park, H. *et al.* A distinct lineage of CD4 T cells regulates tissue inflammation by producing interleukin 17. *Nature Immunol.* **6,** 1133–1141 (2005).
26. Hsieh, C. S., Zheng, Y., Liang, Y., Fontenot, J. D. & Rudensky, A. Y. An intersection between the self-reactive regulatory and nonregulatory T cell receptor repertoires. *Nature Immunol.* **7,** 401–410 (2006).
27. Hsieh, C. S. *et al.* Recognition of the peripheral self by naturally arising $CD25^+$ $CD4^+$ T cell receptors. *Immunity* **21,** 267–277 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research is supported by the NIH, American Diabetes Association and Howard Hughes Medical Institute. R.A.F. is an investigator of the Howard Hughes Medical Institute. Y.Y.W. is supported by a postdoctoral fellowship from the Cancer Research Institute. We thank L. Evangelisti, C. Hughes and J. Stein for assisting with the generation of *FILIG* mice. We are grateful to E. Eynon and L. Zenewicz for critical reading and helpful comments. We also thank F. Manzo and R. Champion for secretarial assistance.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.A.F. (richard.flavell@yale.edu).

LETTERS

Foxp3-dependent programme of regulatory T-cell differentiation

Marc A. Gavin[1]†, Jeffrey P. Rasmussen[1], Jason D. Fontenot[1], Valeria Vasta[2], Vincent C. Manganiello[4], Joseph A. Beavo[2] & Alexander Y. Rudensky[1,3]

Regulatory CD4$^+$ T cells (T_R cells), the development of which is critically dependent on X-linked transcription factor Foxp3 (forkhead box P3), prevent self-destructive immune responses[1]. Despite its important role, molecular and functional features conferred by Foxp3 to T_R precursor cells remain unknown. It has been suggested that Foxp3 expression is required for both survival of T_R precursors as well as their inability to produce interleukin (IL)-2 and independently proliferate after T-cell-receptor engagement, raising the possibility that such 'anergy' and T_R suppressive capacity are intimately linked[2–4]. Here we show, by dissociating Foxp3-dependent features from those induced by the signals preceding and promoting its expression in mice, that the latter signals include several functional and transcriptional hallmarks of T_R cells. Although its function is required for T_R cell suppressor activity, Foxp3 to a large extent amplifies and fixes pre-established molecular features of T_R cells, including anergy and dependence on paracrine IL-2. Furthermore, Foxp3 solidifies T_R cell lineage stability through modification of cell surface and signalling molecules, resulting in adaptation to the signals required to induce and maintain T_R cells. This adaptation includes Foxp3-dependent repression of cyclic nucleotide phosphodiesterase 3B, affecting genes responsible for T_R cell homeostasis.

In males, Foxp3 deficiency results in fatal early-onset systemic autoimmune disease[5]. In heterozygote *Foxp3$^{wt/null}$* females only one-half of T cells harbours the mutant *Foxp3* allele due to random X-chromosome inactivation, whereas autoimmunity is controlled by a normal T_R population expressing the *Foxp3* wild-type allele. Thus, we were able to genetically mark cells actively transcribing a *Foxp3^{null}* allele, yet lacking Foxp3 protein (hereafter called T_{FN} for *Foxp3^{null}*-expressing T cells), through an in-frame insertion of *GFP* into a stop-codon-disrupted *Foxp3* locus (*Foxp3^{gfpko}*) and investigate their features in mice (Fig. 1a; see also Supplementary Figs 1 and 2a). Female *Foxp3$^{gfpko/wt}$* mice were healthy, whereas male *Foxp3^{gfpko}* mice developed the same severity of autoimmunity as *Foxp3* knockout (*Foxp3^{null}*) mice[6], resulting in death at ~4 weeks of age. Thymocyte and peripheral lymphoid organ cellularity did not differ between *Foxp3$^{gfpko/wt}$* and *Foxp3$^{gfp/gfp}$* mice, nor did the proportion of Foxp3$^+$ T_R cells and Foxp3$^-$ CD4$^+$ T cells (data not shown). As our main focus was to characterize T_{FN} cells in healthy *Foxp3$^{gfpko/wt}$* mice, analysis of autoimmune male *Foxp3^{gfpko}* mice is included as Supplementary Fig. 2.

T_{FN} cells constituted ~1–3% of mature CD4$^+$ thymocytes and peripheral CD4$^+$ T cells, indicating that Foxp3 is not required to rescue T_R precursors from negative selection (Fig. 1b, c). This is consistent with a reported abundance of T-cell receptors (TCRs) characteristic of T_R cells in *Foxp3^{null}* mice[7]. As ectopic expression of Foxp3 has been shown to induce a state of hyporesponsiveness in CD4$^+$ T cells[3], we expected T_{FN} cells to appear highly activated, similar to CD25$^+$Foxp3$^-$CD4$^+$ T cells (Fig. 1b), which exhibit characteristics of activated, pro-inflammatory T cells[8]. T_{FN} cells, however, were small and CD62L^{high}, more similar to T_R cells and naive CD4$^+$ cells than to CD25$^+$Foxp3$^-$CD4$^+$ T cells (Fig. 1b, d). T_{FN} cells expressed intermediate levels of CD25, CD44, CTLA4, GITR and ICOS in comparison to T_R cells (Fig. 1b, d; see also Supplementary Fig. 3a). Notably, reduced IL-7R expression, which is considered to be a distinguishing feature of T_R cells, was lowest on T_{FN} cells in comparison to the other T-cell subsets (Fig. 1d). In contrast to T_R cells, T_{FN} cells exhibited negligible proliferative activity *in vivo* based on proliferation-associated antigen Ki67 expression (Fig. 1e).

In vitro, both T_{FN} and T_R cells showed an anergic phenotype, which was reversed in both T_{FN} and T_R cells by provision of IL-2; however, proliferation was more readily restored in T_{FN} cells by limited CD28 co-stimulation (Fig. 1f). Thus, anergy in T_{FN} cells is less stable in comparison to T_R cells. T_{FN} cells were also similar to T_R cells—and distinct from T_N (GFP$^-$CD25$^-$CD4$^+$ T cells from *Foxp3$^{gfp/gfp}$* mice) and CD25$^+$Foxp3$^-$CD4$^+$ T cells—in their inability to generate IL-2 and T-helper 1 and 2 (T_H1 and T_H2) cytokines (Fig. 1g). Thus, IL-2 expression may be blocked transcriptionally during T_R cell differentiation in a Foxp3-independent manner. Lack of IL-2 production and decreased IL-7R make both T_{FN} and T_R cells reliant on exocrine IL-2, but T_{FN} cells are probably less competitive due to intermediate CD25 levels. Indeed, we found that T_{FN} cells were hyper-responsive to elevated IL-2 levels *in vivo* (Supplementary Fig. 7). Together, our analyses show that T_R cell development coincides with Foxp3-independent acquisition of several key T_R cell characteristics.

Unlike T_R cells, some T_{FN} cells produced either IL-10 or IL-17, resembling Tr1 (IL-10-producing Foxp3$^-$CD4$^+$ T cells[9]) and T_H17 cells, respectively (Fig. 1g; see also Supplementary Fig. 2g). Consistent with observed IL-17 production, both T_{FN} and CD25$^+$Foxp3$^-$CD4$^+$ T cells were found to express orphan nuclear receptor RORγt (*Rorc*), which has been shown to both promote IL-17 and block IL-2 production in T_H17 cells[10] (Supplementary Fig. 3b). Notably, *Rorc* transcription was reduced only fourfold in T_R relative to T_{FN} cells but *Il17* transcription was suppressed 7,500-fold, suggesting an ability of Foxp3 to over-ride RORγt activity.

The defining feature of T_R cells is their ability to suppress pro-inflammatory immune function. T_{FN} cells were incapable of suppressing CD4$^+$CD25$^-$ T-cell proliferation *in vitro* (Fig. 2a) and controlling effector T-cell expansion and ensuing splenomegaly and lymphadenopathy after adoptive co-transfer into T-cell-deficient recipient mice (Fig. 2b, c). Spleens and lymph nodes of mice that had received CD45RB^{high}CD25$^-$CD4$^+$ effector T cells with T_{FN} cells were similar in cellularity to those of mice that had received effector T-cells alone, whereas the co-transfer of T_R cells with effector T cells

[1]Departments of Immunology and [2]Pharmacology, and the [3]Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA. [4]National Heart, Lung, and Blood Institute, NIH, 9000 Rockville Pike, Bethesda, Maryland 20892, USA. †Present address: Amgen Corporation, Seattle, Washington 98101, USA.

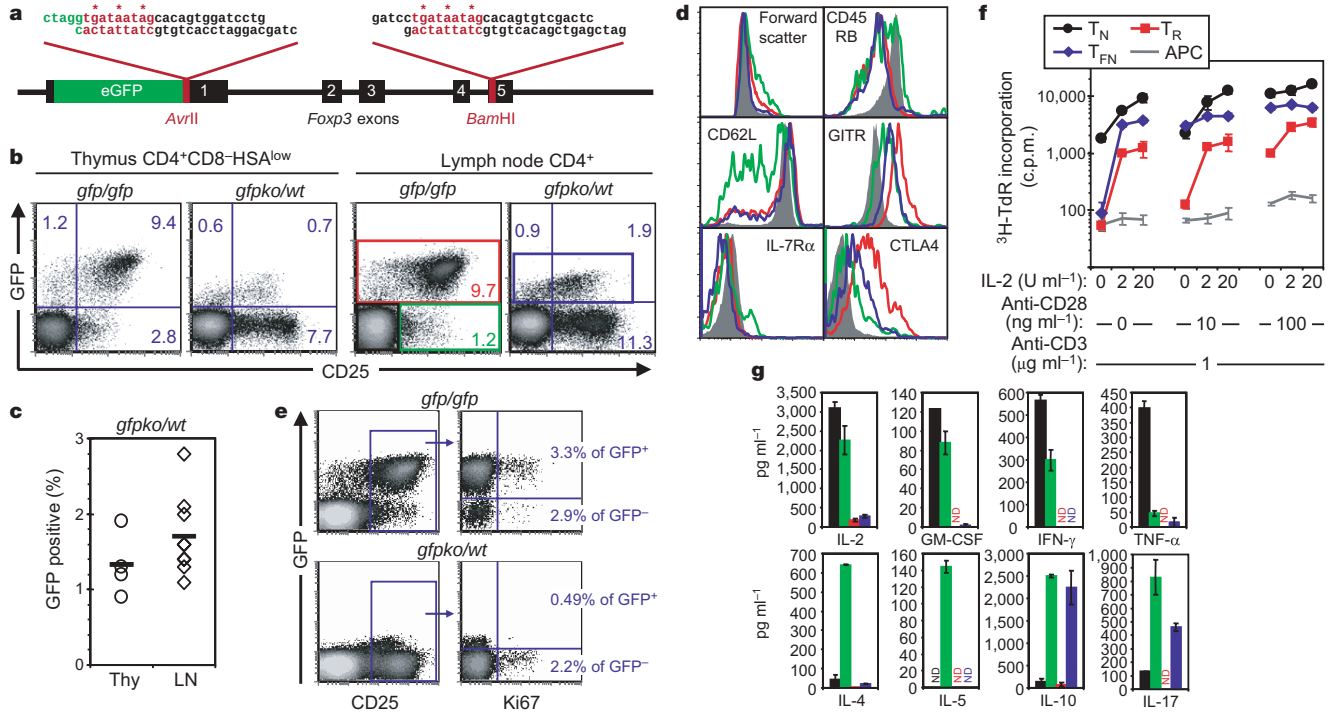


Figure 1 | Phenotype and frequency of cells transcribing the *Foxp3*^{gfpko} locus in *Foxp3*$^{gfpko/wt}$ mice. **a**, We modified our previously described GFP–Foxp3 fusion protein targeting construct[8] by inserting stop codon/frameshift cassettes downstream of GFP and also into *Foxp3* exon 5. In this manner, potentially important regulatory sequences were retained without the possibility that cryptic re-initiation of Foxp3 translation would generate full-length protein. **b**, Flow cytometric analysis of $CD4^+CD8^-HSA^{low}$ thymocytes or $CD4^+$ lymph node cells from female *Foxp3*$^{gfp/gfp}$ or *Foxp3*$^{gfpko/wt}$ mice. **c**, Frequency of GFP^+ cells among either $CD4^+CD8^-HSA^{low}$ thymocytes (Thy, circles) or $CD4^+$ lymph node cells (LN, diamonds) for individual mice. **d**, Flow cytometric analysis of $CD4^+$ lymph node cells for the indicated markers. The four cell subsets are demarcated in **b** as follows: GFP^-CD25^- T_N cells (grey), $GFP^-CD25^+Foxp3^-CD4^+$ T cells (green), GFP^+ T_R cells (red) from *Foxp3*$^{gfp/gfp}$ mice, and GFP^+ T_{FN} cells from *Foxp3*$^{gfpko/wt}$ mice (blue). **e**, Ki67 staining of lymph node cells from the indicated mice. **f**, Proliferation of sorted T cells in response to the indicated reagents. Results are representative of three independent experiments. Data represent the mean and standard deviation for triplicate wells. **g**, Cytokine production by sorted cells from *Foxp3*$^{gfp/gfp}$ (T_N, $CD25^+Foxp3^-CD4^+$, T_R cells) or *Foxp3*$^{gfpko/wt}$ (T_{FN} cells) mice in response to plate-bound anti-CD3 and anti-CD28 assayed by enzyme-linked immunosorbent assay. Results are representative of three separate experiments; data represent mean and maximum/minimum for duplicate wells and colours are as described in **d**.

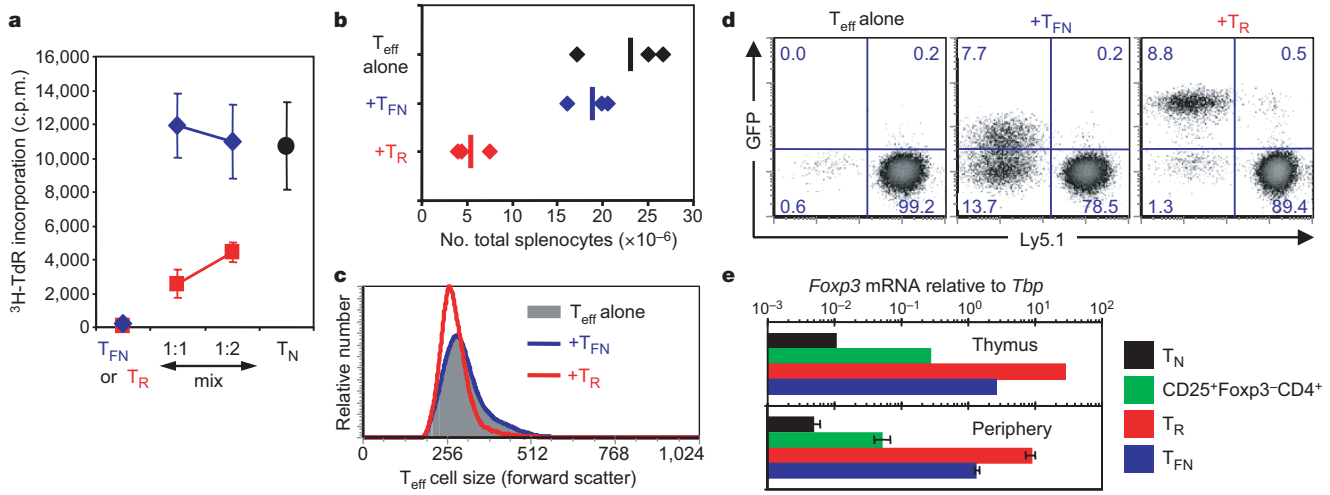


Figure 2 | Foxp3-dependent suppression and lineage stability. **a**, FACS purified T_R or T_{FN} cells were compared for their ability to suppress $CD25^-CD4^+$ T-cell proliferation in the presence of concavalin A and T-cell-depleted splenic antigen presenting cells. Data represent mean and standard deviation for triplicate wells; results are representative of four separate experiments. **b–d**, To examine suppressor activity *in vivo*, T_R or T_{FN} cells were co-transferred with allelically marked effector T cells into lymphopenic animals. Sorted $Ly5.1^+CD45RB^{high}CD25^-CD4^+$ effector T cells (T_{eff}, 3×10^5) were transferred into B6.SCID mice alone or with sorted T_R or T_{FN} cells (10^5). Five weeks after transfer, splenocytes were enumerated (**b**) and lymph node and spleen cells were analysed by flow cytometry. Cell size of the effector T-cell population (**c**) and abundance and GFP expression of the donor T_R or T_{FN} cell populations (**d**) are shown. Data are representative of two independent experiments. **e**, Quantification of *Foxp3* cDNA prepared from the indicated cell populations. For peripheral cells, error bars represent mean and maximum/minimum for biological duplicates.

significantly limited effector T-cell expansion. Similarly, effector T cells were increased in size in the presence of T_{FN} cells, but not T_R cells, likely reflecting increased proliferation (Fig. 2c). Lack of suppression was not due to loss of T_{FN} cells because donor T_{FN} cells competed efficiently with effector T cells during homeostatic expansion, manifested by an increase in their proportion and number (Fig. 2d).

T_{FN} cells also differed from T_R cells in their ability to retain *Foxp3* transcription. Whereas ~90% of donor T_R cells from *Foxp3*$^{gfp/gfp}$ mice retained expression of GFP-tagged Foxp3 protein, two-thirds of the expanded T_{FN} population lost GFP expression (Fig. 2d). Thus, in the course of homeostatic proliferation, *Foxp3* transcription in the absence of Foxp3 function was not efficiently retained. Furthermore, quantification of *Foxp3* messenger RNA in both thymic and peripheral T_{FN} and T_R cells supported the model that Foxp3 promotes its own transcription (Fig. 2e). Although we cannot rule out the possibility that the premature stop codons destabilize *Foxp3*^{gfpko} mRNA, we have found that conditional deletion of *Foxp3* exons 1–5 in mature T_R cells also results in an identical reduction of *Foxp3* mRNA levels[11]. Together, these findings demonstrate that stability of *Foxp3* expression and suppressor activity is dependent upon Foxp3 function rather than coinciding with its expression.

To characterize the Foxp3-dependent transcriptional programme, global gene expression profiling was performed. Automated gene clustering based on similarity in fold-change values between T_N cells and CD25$^+$Foxp3$^-$CD4$^+$ T cells, T_{FN} cells or T_R cells revealed that a distinct set of genes with moderate fold-change values in T_{FN} cells was further amplified in T_R cells. These genes encoded many of the cell surface markers currently used to identify T_R cells, such as CD25, GITR and CTLA4. To incorporate this expression pattern, gene clusters were then hand-curated to generate eight clusters with 'Foxp3-amplified' genes residing in clusters T4 and P4 (Fig. 3a, b; see also Supplementary Fig. 5). For both thymocytes and peripheral cells, Foxp3-dependent genes constituted the largest cluster, reflecting a dominant role for Foxp3 in broadly effecting cellular physiology and phenotype (Fig. 3a, b; clusters T3, P3). Notably, 'Foxp3-dependent' (T3/P3) and 'Foxp3-amplified' genes (T4/P4) most faithfully retained the same expression pattern in thymic and peripheral cells, with Foxp3-amplified genes containing the highest proportion of shared genes (Supplementary Fig. 6a). These findings corroborate our functional studies in elucidating a novel facet of T_R cell differentiation whereby transcriptional and functional characteristics induced concurrently or before *Foxp3* transcription are enforced by Foxp3 to become principal features of T_R cell biology.

To ascribe functional meaning to the defined gene clusters, comparisons with gene annotation databases were performed (Supplementary Fig. 6b, c), revealing that the Foxp3-dependent genes shared between thymic and peripheral data sets were significantly enriched for genes encoding cell surface and extracellular proteins. In contrast, the most enriched functional category for peripheral T_R cells was cell-cycle-associated genes (P6), demonstrating that Foxp3 rescues the inferior proliferative activity of T_{FN} cells in agreement with the analyses of cell size and Ki67 expression (Fig. 1d, e). Also for peripheral cells, the gene cluster specifically regulated by Foxp3 (P3) was enriched for genes involved in intercellular communication.

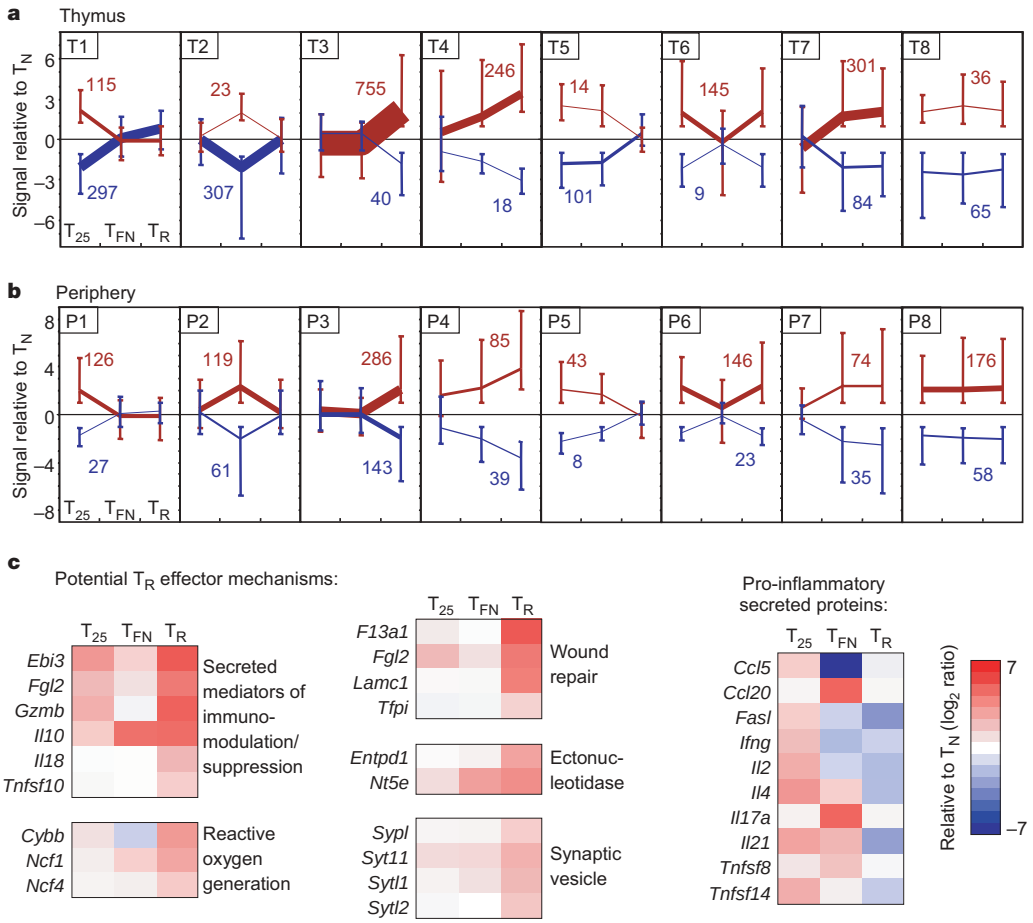


Figure 3 | Foxp3-amplified gene expression constitutes a defining characteristic of thymic and peripheral T_R cells. **a**, **b**, Genes differentially expressed in CD25$^+$Foxp3$^-$CD4$^+$ (T_{25}), T_R and T_{FN} cells relative to T_N cells were identified for both thymic (**a**) and peripheral (**b**) T-cell subsets. For each cluster, values and line thicknesses represent the number of upregulated or downregulated genes, and error bars represent the full range of $\log_2$ expression ratios. **c**, Expression patterns of genes that may have a role, or have been demonstrated to have a role, in T_R cell effector function. Foxp3-mediated repression of genes encoding pro-inflammatory cytokines and chemokines is also shown.

Together, these data suggest that a second major role for Foxp3 is to adapt developing T_R cells by altering how environmental cues are integrated into cellular processes.

Because our experiments showed that suppressor function was Foxp3-dependent, we examined clusters T3 and P3 for genes defining potential T_R effector mechanisms (Fig. 3c). This analysis suggested several such mechanisms, including suppressive soluble factors[12–14], generation of extracellular adenosine[15–18], and release of reactive oxygen[19], and a possible, previously unanticipated role for T_R cells in regulating wound repair.

Among Foxp3-dependent genes, *Pde3b* (cyclic nucleotide phosphodiesterase 3B, cGMP-inhibited) was the most repressed (Fig. 4a, b). As cyclic AMP–protein kinase A pathways are well characterized attenuators of T-cell function[20], reduced PDE3B levels implied cAMP-mediated adaptation in T_R cells to chronic TCR and IL-2 signalling. Notably, naive T cells fail to downregulate PDE3B upon TCR engagement (Fig. 4b), and we found that Foxp3 binds a highly conserved region in the first intron of *Pde3b* (ref. 21). Thus, reduced PDE3B expression represents the first unique marker of T_R cells and may be considered more definitive then Foxp3 itself as it reports Foxp3 function. Hypothesizing that increased expression of PDE3B in T_R cells would be deleterious to T_R cell homeostasis, we introduced PDE3B or a catalytically inactive mutant, PDE3B(H801A)[22], into T_R cells by retroviral gene transfer (Supplementary Fig. 8a, f, g). After transfer into T-cell-deficient recipients, PDE3B-expressing T_R cells were reduced in number by ~60% in comparison to empty vector controls, whereas ectopic PDE3B expression in Foxp3$^-$CD4$^+$ T cells (T_H) had no effect on their homeostasis (Fig. 4c). T_R cells expressing PDE3B(H801A) also exhibited reduced numbers, although to a lesser degree. Loss of PDE3B-transduced T_R cells was due to cell death or inefficient expansion rather than loss of T_R phenotype, because enumeration of expanded donor T_R and T_H cells based on differing Ly5 congenic markers rather than T_R phenotype revealed the same reduction of T_R cell numbers (Fig. 4c).

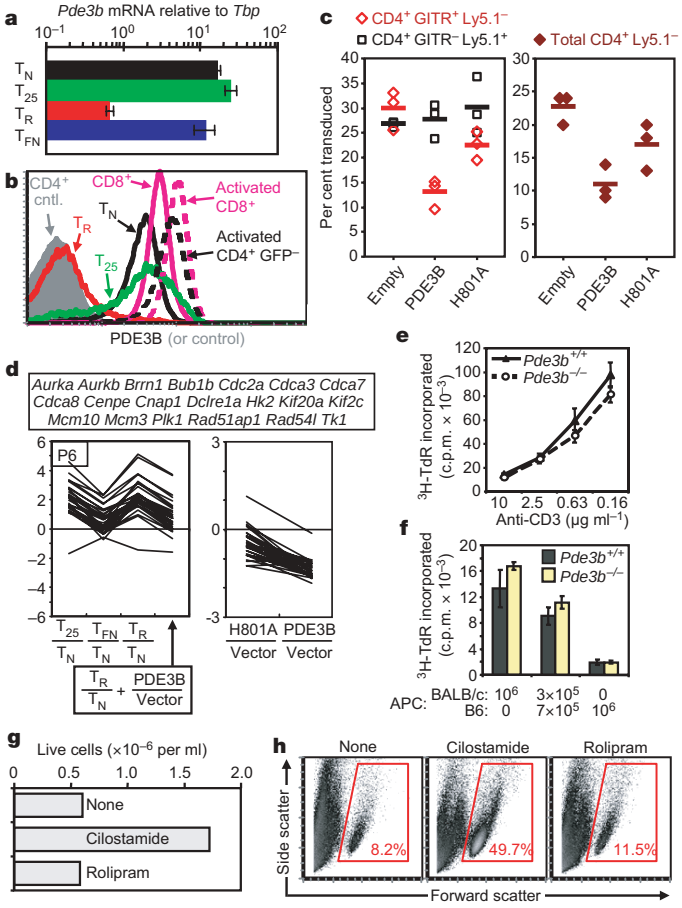


Figure 4 | Reduced PDE3B expression permits normal T_R cell homoeostasis and T_R-specific gene expression. **a**, **b**, *Pde3b* gene expression was analysed by real-time polymerase chain reaction (**a**) and by flow cytometry (**b**). Error bars represent mean and maximum/minimum for biological duplicates. Expression in GFP$^-$CD4$^+$ and CD8$^+$ gated cells after 3 days stimulation of total *Foxp3*$^{gfp/gfp}$ splenocytes with ConA is also shown (dashed lines). T_{25}, CD25$^+$Foxp3$^-$CD4$^+$ T cells. **c**, Frequency in recipient mice of transduced (GFP$^+$) T_R cells (Ly5.1$^-$CD4$^+$GITR$^+$) or T_N cells (Ly5.1$^+$CD4$^+$GITR$^-$), as well as the frequency of transduced Ly5.1$^-$CD4$^+$ cells (see Supplementary Fig. 8a). Results are representative of five separate experiments. **d**, The effect of ectopic PDE3B (PDE3B/Vector) on expression of the listed cell-cycle-associated genes in cluster P6 was added to T_R/T_N values for the same genes (Fig. 3b) and plotted as described (Supplementary Fig. 8c). The effect of PDE3B(H801A) and PDE3B on the same genes is also shown. **e**, **f**, Proliferative responses of *Pde3b*$^{+/+}$ or *Pde3b*$^{-/-}$ CD4$^+$ T cells to anti-CD3 in the presence of irradiated APC (**e**) and to allogeneic (BALB/c) stimulator cells (**f**). Results are representative of two separate experiments; data represent mean and standard deviation for triplicate wells. **g**, **h**, Pre-activated CD4$^+$CD25$^-$ T cells were cultured with IL-2 alone or with the PDE3 inhibitor cilostamide or the PDE4 inhibitor rolipram as described in Supplementary Methods. Viable cell counts (**g**) and percentage live cells (**h**) are shown. Results are representative of three separate experiments.

Transcriptional profiling of recovered cells suggested mechanistic underpinnings for reduced homeostatic fitness of PDE3B-transduced T_R cells, including reduced expression of genes encoding mitochondrial and biosynthetic proteins (Supplementary Fig. 8b). Remarkably, cell cycle genes shared by CD25$^+$Foxp3$^-$CD4$^+$ T cells and T_R cells (cluster P6) reverted to T_{FN}-like expression levels in PDE3B-transduced T_R cells in a manner that was largely dependent on PDE3B catalytic activity (Fig. 4d). This reversion to T_{FN}-like gene expression was also observed for many other Foxp3-dependent genes (Supplementary Fig. 8c). In apparent contrast to the deleterious effect of ectopic PDE3B on T_R cell homeostasis, we found no difference in the proliferative responses of PDE3B-deficient and -sufficient CD4$^+$ T cells[23], and proliferation of wild-type T cells was not altered by PDE3 inhibitors (Fig. 4e, f and data not shown). However, we observed a substantial increase in viability and numbers of pre-activated T_N cells after IL-2-driven expansion in the presence of a PDE3, but not PDE4, inhibitor, suggesting that apoptosis associated with chronic IL-2-induced proliferation may be facilitated by PDE3B activity (Fig. 4g, h). Thus, we have identified *Pde3b* repression as a central component of Foxp3-dependent T_R cell maintenance, as its re-expression resulted in a reduction of biosynthetic processes, the attenuation of Foxp3-dependent proliferative fitness and the loss of some Foxp3-dependent gene expression.

Our study suggests that rather than initiating a *de novo* developmental programme in self-reactive T cells, Foxp3 takes advantage of preceding and coincidental features of T_R precursor cells probably facilitated by TCR signalling. This is accomplished through consolidation of the state of anergy and paracrine IL-2 dependence. Furthermore, Foxp3 amplifies and stabilizes expression of genes encoding cell surface or secreted molecules—like Fgl2, CD73, CD39, TRAIL or CTLA4—normally elaborated by conventional T cells upon TCR stimulation and capable of negative feedback regulation of T-cell activation '*in trans*'. At the same time, Foxp3 enforces repression of TCR-activation-dependent immune response effector cytokines including IL-4, interferon-γ, tumour-necrosis factor-α, IL-17 and IL-21. These observations might explain failure to identify a single non-redundant mechanism of suppression mediated by T_R cells. Finally, Foxp3 alters how T_R cells respond to environmental cues by modulating cell surface and signalling molecules to promote both T_R homeostasis and lineage stability. Essential adaptation to these signals results from Foxp3-dependent downregulation of PDE3B to support normal homeostasis and metabolic function and maintain a part of the T_R cell transcriptional programme. Our results suggest that, in resemblance of evolutionary processes, a lineage commitment factor may act in both an opportunistic and

adaptive fashion during cellular differentiation by amplifying beneficial and correcting disabling features of precursor cells.

METHODS

Mice. B6.SCID, TCR$\alpha^{-/-}$ and TCR$\beta\delta^{-/-}$ mice were obtained from The Jackson Laboratory, and C57BL/6 mice were from the Charles River Breeding Laboratories. All mice were maintained at the University of Washington specific pathogen-free facility. Generation of *Foxp3*^{gfpko} is described in Supplementary Methods.

Flow cytometry and cell sorting. Staining and sorting of cells and intracellular staining for Ki67 and PDE3B are described in Supplementary Methods.

***In vitro* assays.** T-cell stimulation, suppression and cytokine release assays were performed according to conventional methods as described in Supplementary Methods. To examine cell survival in the presence of PDE inhibitors, $CD25^{-}CD4^{+}$ T cells were stimulated with plate-bound anti-CD3 and anti-CD28 antibodies. On day 2 cells were removed from antibody-coated plates and on day 4 cells were washed and plated at 2×10^5 cells ml^{-1} with IL-2 (100 U ml^{-1}) or IL-2 plus cilostamide or rolipram (each at 5 mM). On day 7 cells were split 1:4 into the same conditions and on day 13 cells were counted and analysed by flow cytometry.

Gene expression analysis. Gene expression was determined with Affymetrix mouse 430 2.0 microarrays and clusters were generated by Pearson squared *K*-means clustering with MultiExperiment Viewer (TIGR) and by hand, as described (Supplementary Methods).

PDE3B expression. Mixtures of transduced and non-transduced $CD25^{+}CD4^{+}$ cells from B6 mice and $CD25^{-}CD4^{+}$ cells from B6.SJL mice ($Ly5.1^{+}$) were transferred into TCR$\alpha^{-/-}$ recipients. Three weeks after transfer cells were analysed by flow cytometry and isolated by fluorescence-activated cell sorting (FACS) for gene expression analysis (Supplementary Methods and Supplementary Fig. 8).

Received 17 November; accepted 20 December 2006.
Published online 14 January 2007.

1. Fontenot, J. D. & Rudensky, A. Y. A well adapted regulatory contrivance: regulatory T cell development and the forkhead family transcription factor Foxp3. *Nature Immunol.* **6,** 331–337 (2005).
2. van Santen, H. M., Benoist, C. & Mathis, D. Number of T reg cells that differentiate does not increase upon encounter of agonist ligand on thymic epithelial cells. *J. Exp. Med.* **200,** 1221–1230 (2004).
3. Hori, S., Nomura, T. & Sakaguchi, S. Control of regulatory T cell development by the transcription factor Foxp3. *Science* **299,** 1057–1061 (2003).
4. Khattri, R., Cox, T., Yasayko, S. A. & Ramsdell, F. An essential role for Scurfin in $CD4^{+}CD25^{+}$ T regulatory cells. *Nature Immunol.* **4,** 337–342 (2003).
5. Wildin, R. S. & Freitas, A. IPEX and FOXP3: clinical and research perspectives. *J. Autoimmun.* **25** (suppl.), 56–62 (2005).
6. Fontenot, J. D., Gavin, M. A. & Rudensky, A. Y. Foxp3 programs the development and function of $CD4^{+}CD25^{+}$ regulatory T cells. *Nature Immunol.* **4,** 330–336 (2003).
7. Hsieh, C. S., Zheng, Y., Liang, Y., Fontenot, J. D. & Rudensky, A. Y. An intersection between the self-reactive regulatory and nonregulatory T cell receptor repertoires. *Nature Immunol.* **7,** 401–410 (2006).
8. Fontenot, J. D. *et al.* Regulatory T cell lineage specification by the forkhead transcription factor foxp3. *Immunity* **22,** 329–341 (2005).
9. Roncarolo, M. G. *et al.* Interleukin-10-secreting type 1 regulatory T cells in rodents and humans. *Immunol. Rev.* **212,** 28–50 (2006).
10. Ivanov, I. I. *et al.* The orphan nuclear receptor RORγt directs the differentiation program of proinflammatory IL-17(+) T helper cells. *Cell* **126,** 1121–1133 (2006).
11. Williams, L. M. & Rudensky, A. Y. Maintenance of the Foxp3 dependent developmental program in mature regulatory T cells requires continued expression of Foxp3. *Nature Immunol.* (in the press).
12. Chan, C. W. *et al.* Soluble fibrinogen-like protein 2/fibroleukin exhibits immunosuppressive properties: suppressing T cell proliferation and inhibiting maturation of bone marrow-derived dendritic cells. *J. Immunol.* **170,** 4036–4044 (2003).
13. Stumhofer, J. S. *et al.* Interleukin 27 negatively regulates the development of interleukin 17-producing T helper cells during chronic inflammation of the central nervous system. *Nature Immunol.* **7,** 937–945 (2006).
14. Batten, M. *et al.* Interleukin 27 limits autoimmune encephalomyelitis by suppressing the development of interleukin 17-producing T cells. *Nature Immunol.* **7,** 929–936 (2006).
15. Yegutkin, G. G., Henttinen, T., Samburski, S. S., Spychala, J. & Jalkanen, S. The evidence for two opposite, ATP-generating and ATP-consuming, extracellular pathways on endothelial and lymphoid cells. *Biochem. J.* **367,** 121–128 (2002).
16. Yegutkin, G. G., Samburski, S. S., Jalkanen, S. & Novak, I. ATP-consuming and ATP-generating enzymes secreted by pancreas. *J. Biol. Chem.* **281,** 29441–29447 (2006).
17. Ohta, A. & Sitkovsky, M. Role of G-protein-coupled adenosine receptors in downregulation of inflammation and protection from tissue damage. *Nature* **414,** 916–920 (2001).
18. Erdmann, A. A. *et al.* Activation of Th1 and Tc1 cell adenosine A2A receptors directly inhibits IL-2 secretion *in vitro* and IL-2-driven expansion *in vivo*. *Blood* **105,** 4707–4714 (2005).
19. Sheppard, F. R. *et al.* Structural organization of the neutrophil NADPH oxidase: phosphorylation and translocation during priming and activation. *J. Leukoc. Biol.* **78,** 1025–1042 (2005).
20. Torgersen, K. M., Vang, T., Abrahamsen, H., Yaqub, S. & Tasken, K. Molecular mechanisms for protein kinase A-mediated modulation of immune function. *Cell. Signal.* **14,** 1–9 (2002).
21. Zheng, Y. *et al.* Genome-wide analysis of Foxp3 target genes in developing and mature regulatory T cells. *Nature* doi:10.1038/nature05563 (in the press).
22. Zhang, W. & Colman, R. W. Conserved amino acids in metal-binding motifs of PDE3A are involved in substrate and inhibitor binding. *Blood* **95,** 3380–3386 (2000).
23. Choi, Y. H. *et al.* Alterations in regulation of energy homeostasis in cyclic nucleotide phosphodiesterase 3B-null mice. *J. Clin. Invest.* **116,** 3240–3251 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to all members of the Rudensky laboratory for discussions. We also thank K. Forbush, L. Karpik and T. Chu for help. This work was supported by grants from the Arthritis Foundation and the Leukemia and Lymphoma Society (M.A.G.) and grants from the National Institutes of Health (A.Y.R.). A.Y.R. is a Howard Hughes Medical Institute Investigator.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.Y.R. (aruden@u.washington.edu).

LETTERS

Dll4 signalling through Notch1 regulates formation of tip cells during angiogenesis

Mats Hellström[1,2], Li-Kun Phng[4], Jennifer J. Hofmann[5], Elisabet Wallgard[1,2], Leigh Coultas[6], Per Lindblom[4]†, Jackelyn Alva[5], Ann-Katrin Nilsson[1]†, Linda Karlsson[1]†, Nicholas Gaiano[7], Keejung Yoon[7], Janet Rossant[6], M. Luisa Iruela-Arispe[5], Mattias Kalén[1,2]*, Holger Gerhardt[4]* & Christer Betsholtz[2,3]*

In sprouting angiogenesis, specialized endothelial tip cells lead the outgrowth of blood-vessel sprouts towards gradients of vascular endothelial growth factor (VEGF)-A[1,2]. VEGF-A is also essential for the induction of endothelial tip cells[2], but it is not known how single tip cells are selected to lead each vessel sprout, and how tip-cell numbers are determined. Here we present evidence that delta-like 4 (Dll4)–Notch1 signalling regulates the formation of appropriate numbers of tip cells to control vessel sprouting and branching in the mouse retina. We show that inhibition of Notch signalling using γ-secretase inhibitors, genetic inactivation of one allele of the endothelial Notch ligand *Dll4*, or endothelial-specific genetic deletion of *Notch1*, all promote increased numbers of tip cells. Conversely, activation of Notch by a soluble jagged1 peptide leads to fewer tip cells and vessel branches. Dll4 and reporters of Notch signalling are distributed in a mosaic pattern among endothelial cells of actively sprouting retinal vessels. At this location, Notch1-deleted endothelial cells preferentially assume tip-cell characteristics. Together, our results suggest that Dll4–Notch1 signalling between the endothelial cells within the angiogenic sprout serves to restrict tip-cell formation in response to VEGF, thereby establishing the adequate ratio between tip and stalk cells required for correct sprouting and branching patterns. This model offers an explanation for the dose-dependency and haploinsufficiency of the *Dll4* gene[3–5], and indicates that modulators of Dll4 or Notch signalling, such as γ-secretase inhibitors developed for Alzheimer's disease, might find usage as pharmacological regulators of angiogenesis.

Sprouting angiogenesis shows morphological and molecular similarities to epithelial tubulogenesis[6] and axon guidance[7]. Through dynamic filopodial protrusions, the sprout tip cells migrate, sense and respond to guidance cues provided by soluble, cell- or matrix-bound ligands, such as VEGF(s), netrin(s) and semaphorin(s)[2,8,9]. During tracheal sprouting in *Drosophila melanogaster*, Notch signalling participates in the selection of cells involved in branch fusion and

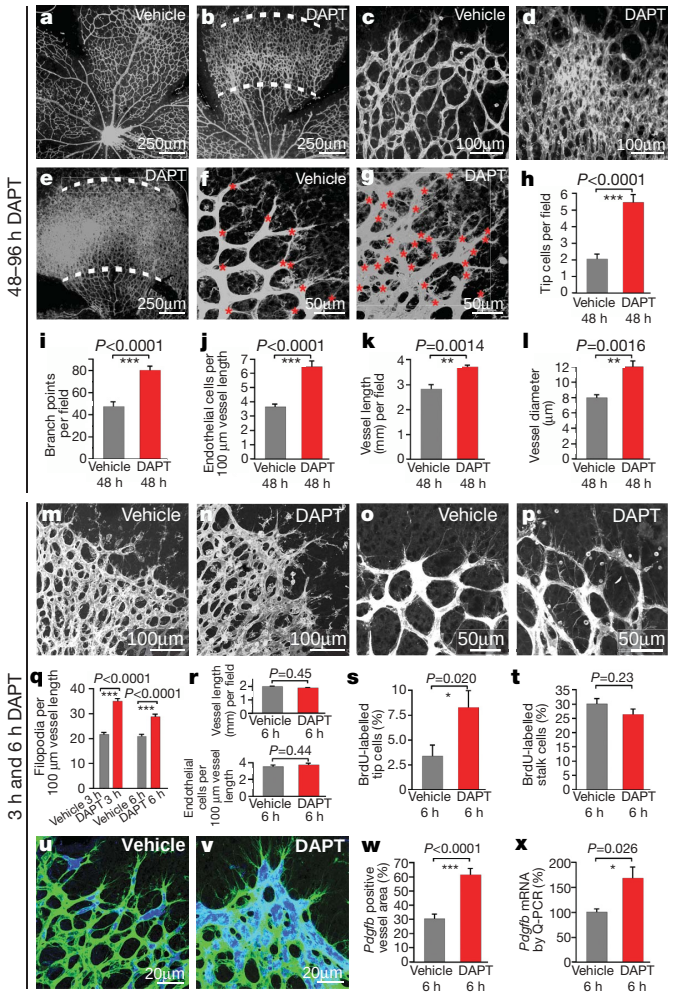


Figure 1 | γ-Secretase inhibitor experiments *in vivo*. Isolectin B4 (white in **a**–**g**, **m**, **n**; green in **u**, **v**) or endomucin staining (white in **o**, **p**) of whole-mounted P4–6 retinas. DAPT treatment lasted for 48 h (**a**–**d**, **f**–**l**), 96 h (**e**), 6 h (**m**–**p**, **r**–**x**) or 3 h (**q**). The affected region (between dashed lines in **b** and **e**) correlates with length of treatment period. **c**, **d**, **f**, **g**, Views of vessel plexus margins. Note the DAPT-induced increase in filopodial protrusions (red asterisks in **f**, **g**). **h**–**l**, Quantifications of densities of tip cells (**h**), branch points (**i**), endothelial cells (**j**), as well as length (**k**) and diameter (**l**) of vessels. **m**–**p**, DAPT-induced increase in filopodia (**o**, **p**), vessel branching (**m**, **n**) and *pdgfb* mRNA expression (blue in **u**, **v**) at the plexus margin. **q**–**t**, Quantifications of the densities of filopodia (**q**), vessels (**r**, top) and endothelial cells (**r**, bottom), as well as of BrdU labelling in tip (**s**) and stalk (**t**) cells. **w**, **x**, Proportion of *pdgfb*-positive endothelium (**w**) and levels of *pdgfb* mRNA (**x**). All error bars represent s.e.m.

[1]AngioGenetics Sweden AB, Scheeles väg 2, SE-171 77 Stockholm, Sweden. [2]Department of Medical Biochemistry and Biophysics, Division of Matrix Biology, and [3]Department of Medicine, Karolinska Institutet, SE 171 77 Stockholm, Sweden. [4]Vascular Biology Laboratory, London Research Institute, Cancer Research UK, London WC2A 3PX, UK. [5]Department of Molecular, Cell and Developmental Biology and Molecular Biology Institute, UCLA, Los Angeles, California 90095, USA. [6]Program for Developmental Biology, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. [7]Neurology, Neuroscience and Oncology, Institute for Cell Engineering, Johns Hopkins University, Baltimore, Maryland 21205, USA.
†Present addresses: Molecular Toxicology, Safety Assessment, AstraZeneca R&D, SE-151 85 Södertälje, Sweden (P.L.); Stem Cell Center, BMC B10, Klinikg. 26, Lund University, SE-221 84 Lund, Sweden (A.-K.N.); Department of Physiology, Göteborg University, P.O. Box 434, SE-405 30 Göteborg, Sweden (L.K.).
*These authors contributed equally to this work.

terminal ramification[10–12]. Because many Notch pathway components are expressed and have critical but poorly defined roles during vascular development[13], we investigated whether Notch signalling might have a specific function in tip-cell formation during sprouting angiogenesis.

To study the importance of Notch during sprouting angiogenesis *in vivo*, we focused on developmental retinal angiogenesis during the first post-natal week in mice. Here, systemic administration of the γ-secretase inhibitors (GSIs) DAPT (*N*-[*N*-(3,5-difluorophenacetyl)-L-alanyl]-*S*-phenylglycine *t*-butylester) or compound X (*S*-3-[*N*′-(3,5-difluorophenyl-α-hydroxy-acetyl)-L-alanilyl]amino-2,3-dihydro-1-methyl-5-phenyl-1H-1,4-benzodiazepin-2-one), which potently inhibit Notch receptor cleavage and signalling[14,15], led to increased vascular density and vessel diameter associated with elevated numbers of endothelial cells and vascular sprouts in the peripheral part of the developing retinal vascular plexus (Fig. 1a–l). DAPT and compound X also increased filopodial protrusions from endothelial cells both at the sprouting front of the plexus, where filopodia are normally prevalent as they extend from the tip cells[2] (Fig. 1f–h), and proximally in the plexus where filopodial protrusions are normally rare (data not shown). JLK-6 (7-amino-4-chloro-3-methoxyisocoumarin), a GSI that does not inhibit Notch, lacked any of these effects (Supplementary Fig. 1). Prolonged treatment with the Notch-inhibitory GSIs augmented the severity of the vessel abnormalities, leading, in the most extreme cases, to vessel coalescence into large, flat sinuses (Fig. 1e). A similar increase in sprout density and change in vascular patterning was observed when DAPT was administered in conjunction with pathological angiogenesis induced in a mouse model of retinopathy of prematurity (ROP) (Supplementary Fig. 2).

The width of the affected zone correlated with the expected plexus extension during the period of treatment, suggesting that GSIs affect developing (distal), but not already formed (proximal), vessels (Fig. 1b, e). Accordingly, short-term (3–6 h) DAPT exposure led to increased filopodial protrusions and vessel diameter only at the margin of the plexus (Fig. 1m–q) and lacked effects on endothelial cell number and vascular density (Fig. 1r). DAPT exposure for 6 h also increased the frequency of endothelial cells positive for the tip-cell marker platelet-derived growth factor-B (*Pdgfb*)[2] at the plexus margin (Fig. 1u–x). Thus, both morphological and gene expression changes suggested that Notch-inhibitory GSIs raise tip-cell numbers, leading to excessive vascular sprouting and fusion. To address whether proliferation might underlie the increase in tip cells, we performed 5-bromodeoxyuridine (BrdU) labelling experiments. DAPT exposure for 6 h did not affect BrdU labelling of stalk cells (Fig. 1t) but raised the proportion of BrdU-positive tip cells (Fig. 1s and Supplementary Fig. 3). This proliferative response in tip cells is, however, too small to explain the much

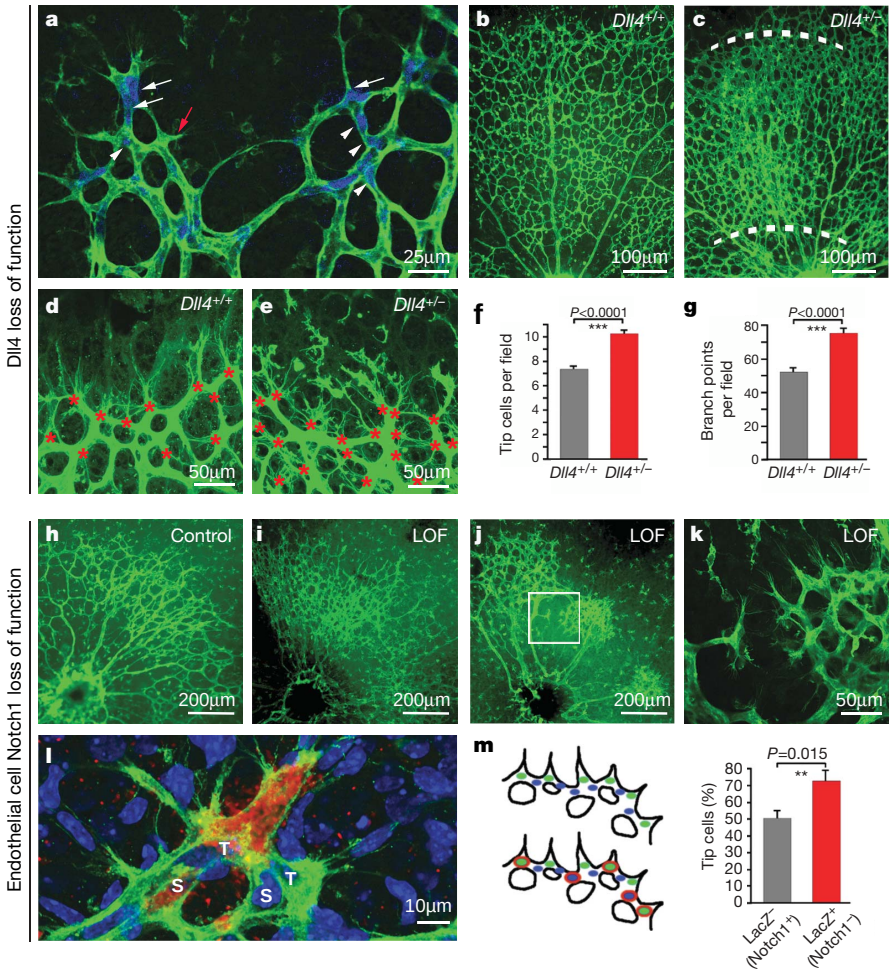


Figure 2 | Dll4 and Notch1 loss-of-function analyses *in vivo*. Isolectin B4 staining (green) of whole-mounted retinas. **a**, mRNA *in situ* hybridization for *Dll4* (blue) on wild-type P5 retina. *Dll4*-positive tip cells (white arrows) and non-tip cells, (white arrowheads), as well as *Dll4*-negative tip cells (red arrow), are indicated. **b**–**g**, Increased vessel density (**b**, **c**, **g**) and filopodia protrusions (**d**–**f**) (red asterisks) in $Dll4^{+/-}$ retinas at P4. **h**–**k**, P5 retinas from control (**h**) and two tamoxifen-treated (LOF) VEcad-$CreER^{T2}$/R26R/$Notch1^{floxed/floxed}$ mice with a different degree of vascular abnormality (**i**, **j**; the box in **j** is magnified in **k**). **l**, Example of an image used to score LacZ-positive (red and green) and LacZ-negative (green) tip cells (T) and stalk cells (S). Nuclei (blue) are stained by TOPRO-3. **m**, Schematic illustration of the scoring principle: retinas with mosaic recombination were analysed at the sprouting plexus margin for tip cells (green) and stalk cells (blue) that were either LacZ positive (red circle) or LacZ negative (no red circle). The graph shows the percentage of tip cells among LacZ-positive and LacZ-negative cells. All error bars represent s.e.m.

larger increases in filopodia extension (Fig. 1q) and *Pdgfb* expression (Fig. 1u–x) observed after short-term DAPT administration.

The Notch-inhibitory activity of DAPT was confirmed by the downregulated retinal messenger RNA levels of the known Notch target gene Notch-regulated ankyrin-repeat protein (*Nrarp*) after 6 h (Supplementary Fig. 4) and an expected increase in the number of intestinal Goblet cells after 4 days[16] (Supplementary Fig. 5). However, because γ-secretase has substrates in addition to Notch that may be involved in angiogenesis[17], we sought specific genetic evidence for a role of Notch signalling in retinal angiogenesis. The restricted vascular expression of the Notch ligand *Dll4* (ref. 18) and the previously described vascular defects of *Dll4* mouse mutants prompted us to study retinal angiogenesis in $Dll4^{+/-}$ mice[3–5]. First, we confirmed the previously reported expression of *Dll4* mRNA in arteries[19] and in a proportion of the tip cells[18], but noticed expression also in a proportion of the endothelial stalk at the sprouting front (Fig. 2a). $Dll4^{+/-}$ mice displayed a similar retinal phenotype compared to long-term GSI-treated pups (Fig. 2b, c). Increased filopodial protrusions, vessel branching and numbers of *Pdgfb*-expressing cells were observed (Fig. 2d–g and Supplementary Fig. 6), consistent with an increased number of tip cells. $Dll4^{+/-}$ retinas displayed morphologically distinguishable arteries and veins (Fig. 2b, c), and normal mural cell abundance and distribution, as revealed by α-smooth muscle actin- and chondroitin sulphate proteoglycan NG2 staining (Supplementary Fig. 7). Thus, whereas previous work demonstrated a role for Notch in arterial specification of both endothelial and mural cells[20,21], the phenotypes observed in GSI-treated and $Dll4^{+/-}$ retinas appeared to reflect mainly an increased formation of tip cells during angiogenic sprouting.

The severe vascular abnormalities and embryonic lethality of both ubiquitous[22] and endothelium-specific[23] Notch1 homozygous null mutants suggest a critical vascular function for Notch1 but precludes postnatal analysis of these mice. To study the role of endothelial Notch1 in retinal angiogenesis we therefore used VEcad-CreER^{T2}/Notch1$^{floxed/floxed}$ mice, in which tamoxifen-inducible Cre enzyme in endothelial cells[24] deletes critical *Notch1* sequences flanked by *loxP* sites[25]. We first analysed the efficiency of tamoxifen-induced recombination in VEcad-CreER^{T2}/R26R retinas after drug administration at postnatal days (P)0–4 and subsequent reporter (LacZ) staining at P5. A mosaic endothelium-specific LacZ expression was observed, ranging between 10–30% in different individuals (Supplementary Fig. 8). The same tamoxifen administration regimen for VEcad-CreER^{T2}/R26R/Notch1$^{floxed/floxed}$ mice resulted in regionally increased sprout and filopodia densities in the retinal vasculature (Fig. 2h–k). We took advantage of individual VEcad-CreER^{T2}/R26R/Notch1$^{floxed/floxed}$ mice that exhibited the lowest degree of recombination resulting in a few scattered LacZ-positive endothelial cells in an otherwise normal retina. Analysis of endothelial cells at the peripheral sprouting front (Fig. 2l) of those mice revealed that a high proportion (73%) of the LacZ-positive cells displayed morphological tip-cell characteristics (Fig. 2m). In contrast, LacZ-negative cells were distributed evenly among tip and stalk cells (50:50) (Fig. 2m). The correlation between Cre expression (and thereby Notch1 inactivation) and acquirement of tip-cell

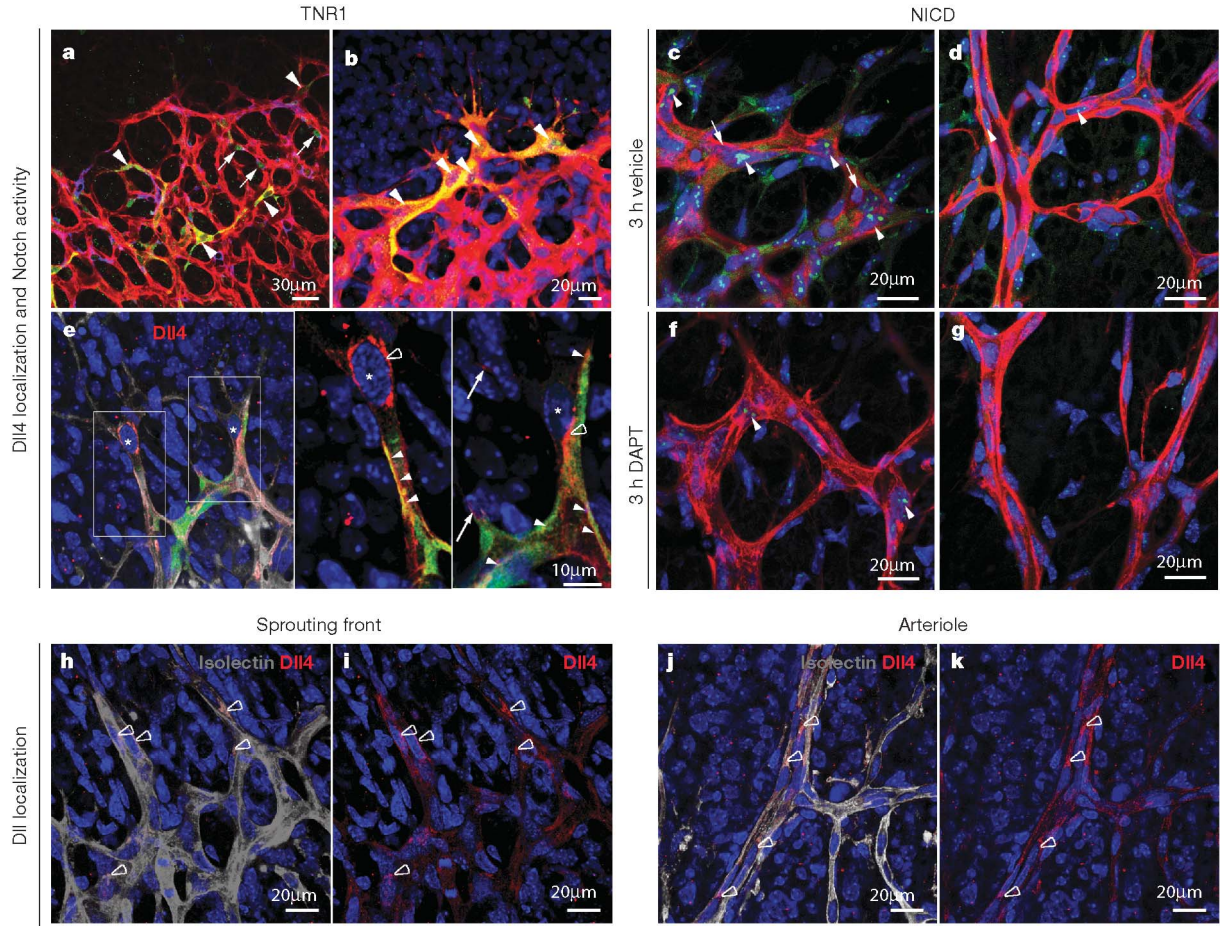


Figure 3 | Endothelial Dll4–Notch signalling *in vivo*. **a–k**, Isolectin B4 (red in **a–d**, **f**, **g**; grey in **e**, **h**, **j**) NG2 (blue in **a**) and DAPI (blue in **b–k**) staining of whole-mounted P4–6 retinas. **a**, **b**, **e**, Notch-reporter activity in TNR1 mouse (GFP; green/yellow) observed in a proportion of endothelial cells localized at the sprouting front (arrowheads). Open arrowheads in **e** indicate typical intracellular Dll4 protein (red) localization in tip cells (asterisks). Left and right boxes in the left panel of **e** are magnified in the middle and right panels, respectively. Filled arrowheads indicate typical membrane Dll4 localization observed in stalk cells and in the plexus. Arrows indicate occasional filopodial Dll4. **c**, **d**, **f**, **g**, Notch1-ICD-positive (turquoise) endothelial nuclei (arrowheads) were abundant at the vascular front (**c**) but scarce in the mature plexus (**d**) and were eliminated at both sites by 3 h of DAPT treatment (**f**, **g**). **h–k**, Internalization of Dll4 (open arrowheads) was detected in tip and stalk cells at the vascular front and in arterioles, but not in other endothelial cells.

characteristics suggests that Notch1 suppresses the tip-cell phenotype in a cell-autonomous manner. This chimaeric analysis also argues against systemic or secondary effects of endothelial Notch1 inhibition on tip-cell formation in the retina.

To demonstrate Notch signalling in sprouting retinal endothelial cells *in vivo*, we used a transgenic Notch-reporter mouse (hereafter called TNR1), which carries a CBF-1 response element and a minimal SV40 promoter followed by an enhanced green fluorescent protein (GFP) sequence[26]. We found GFP expression broadly in the endothelium at the vascular front, and a weak signal in some non-vascular retinal cells, consistent with Notch receptor expression and signalling in several retinal cell types[18]. However, conspicuous GFP expression was only seen in endothelial cells at the sprouting front, distributing in a mosaic pattern involving both tip and stalk cells. Thus, the most active Notch signalling was mainly observed in the vascular region affected by Notch inhibition (Fig. 3a, b). Furthermore, strongly GFP-positive endothelial cells intermingled with Dll4-expressing endothelial cells at the sprouting front (Fig. 3e). Some Dll4-positive cells exhibited intracellular accumulation of Dll4 staining, suggesting previous engagement in Notch activation (Fig. 3e)[27]. Almost all tip cells (48 out of 50 analysed) showed intracellular accumulation of Dll4 staining. Notably, many stalk cells immediately in contact with the tip cells (33 out of 58) also showed intracellular accumulations of Dll4 staining, indicating active bi-directional Dll4 signalling between tip cells and stalk cells at the sprouting front (Fig. 3h, i). Dll4 accumulation was similarly observed in arterioles, a known site of Dll4–Notch signalling (Fig. 3j, k), but endothelial cells in the plexus behind the growing front or adjacent to arterioles showed no or minor intracellular accumulation of Dll4. To demonstrate Notch1 cleavage *in vivo* we used an antibody against Notch1 intracellular domain (ICD). This labelled nuclei of vascular cells, astrocytes and microglia, in agreement with the reported retinal Notch1 expression pattern[18]. As with Notch-reporter activity in TNR1 mice, Notch1-ICD was readily detected in endothelial cells at the sprouting front (Fig. 3c, arrowheads), but rarely and weakly at more proximal locations in the retinal plexus (Fig. 3d, arrowheads). The specificity of the Notch1-ICD signal was shown by its downregulation after DAPT administration (Fig. 3f, g, arrowheads in f show residual signal). Thus, both Notch reporter in TNR1 mice and Notch1-ICD were preferentially located in endothelial cells at sites exposed to VEGF-A[2] (that is, where selection of tip cells occurs).

Previous work showed that tip-cell filopodial extension in the developing retina depends on astrocyte-derived VEGF-A acting directly on tip cells[2]. We therefore investigated whether Notch inhibition affected astrocytes or VEGF-A expression. Glial fibrillary acidic protein (GFAP) staining revealed a normal astrocyte network ahead of the plexus front in GSI-treated and $Dll4^{+/-}$ mice (Supplementary Fig. 9). Retinal VEGF-A mRNA and protein levels as well as the relative levels of the VEGF-A splice isoforms were unchanged in DAPT-treated retinas (Supplementary Fig. 10). *In situ* hybridization also showed that the retinal pattern of VEGF-A expression was normal after 6 h of GSI treatment but shifted proximally to the areas of dense and hyper-fused vasculature after long-term DAPT treatment and in $Dll4^{+/-}$ retinas (Supplementary Fig. 10). Thus, whereas the central retinal vascular abnormalities observed in $Dll4^{+/-}$ mice and after long-term DAPT treatment might, in part, result from changes in the pattern of VEGF-A expression, the short-term effects of GSIs on tip-cell formation at the

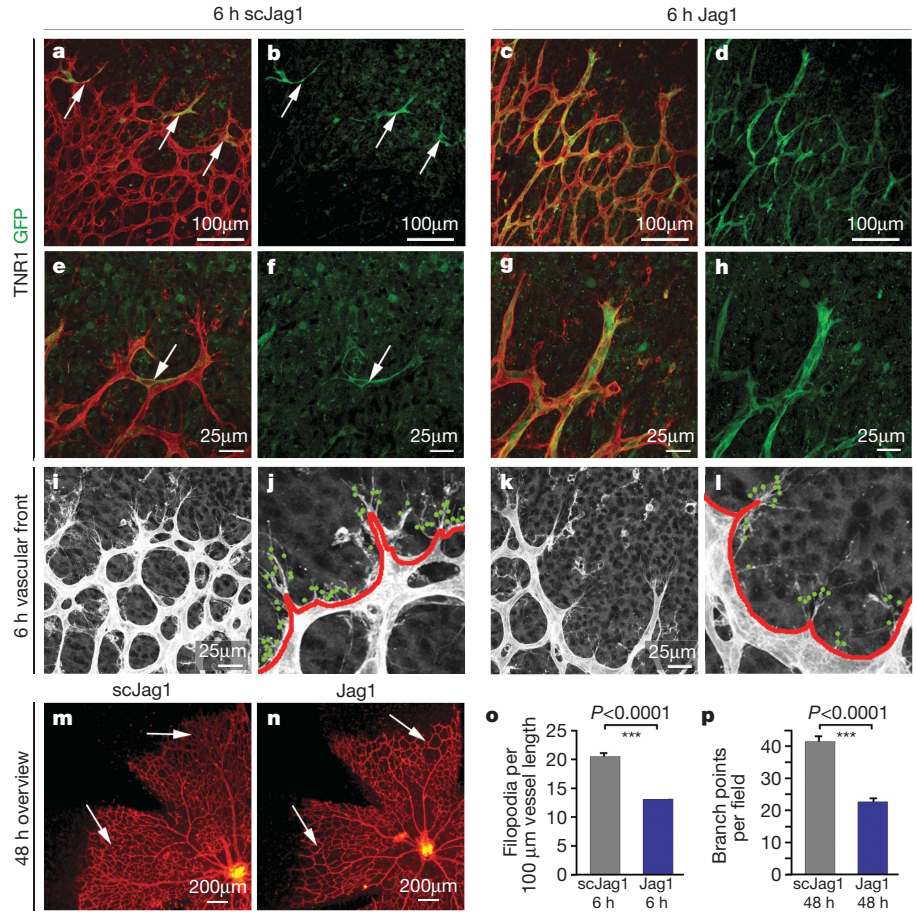


Figure 4 | Notch gain-of-function analysis *in vivo*. Isolectin B4 (red in **a**, **c**, **e**, **g**, **m**, **n**; white in **i**–**l**) staining of whole-mounted P4–5 retinas. **a**–**h**, Broad induction of Notch reporter in TNR1 mouse (GFP; green/yellow) by 6 h treatment with Jag1 (**c**, **d**, **g**, **h**) but not a scrambled peptide (scJag1; **a**, **b**, **e**, **f**; arrows indicate scattered Notch-reporter activity at the front). **i**–**l**, **o**, Six hours of Jag1 treatment reduced tip cells and filopodia at the sprouting front. **j**, **l**, **o**, Quantification of filopodia (green dots) per vessel length (red line) after scJag1 (**j**) and Jag1 (**l**) treatment. **j** and **l** are magnifications of **i** and **k**. Forty-eight hours of Jag1 treatment reduced retinal vessel density in the peripheral vascular plexus (arrows in **m**, **n**; quantification in **p**). All error bars represent s.e.m.

plexus margin are unlikely to be caused by changes in VEGF-A. Short term (6 h) DAPT treatment also did not lead to a significant change in retinal expression of mRNAs encoding the hypoxia marker solute carrier family 2 member 1 (*Slc2a1*) (Supplementary Fig. 10).

Our data suggest that Dll4–Notch1 signalling within the endothelial cell population serves to suppress the tip-cell phenotype. To test this hypothesis in a direct gain-of-function experiment, we used a synthetic peptide corresponding to the δ/serrate/Lag-2 domain of jagged1 (Jag1) that has previously reported Notch agonistic activity[28]. The ability of Jag1 peptide to activate Notch signalling *in vivo* was confirmed in TNR1 mice: systemic administration of Jag1 peptide, but not a scrambled control peptide, induced widespread GFP expression in retinal endothelial cells (Fig. 4a–h; compare panel b to d). It also induced a 35% decrease in filopodia density at the vascular front after 6 h (Fig. 4i–l, o). Repeated Jag1 peptide injections also led to a 45% decrease in vessel density (Fig. 4m, n, p); that is, the opposite effects compared to the different abovementioned situations of Notch inhibition.

The competence to become a tip cell is broad, perhaps ubiquitous, within the endothelial population[2]. VEGF-A gradients and point sources occur in tissues as a result of cell-type-specific expression and matrix binding of VEGF-A, and have been shown to be essential for vascular patterning[1,29]. Although the endothelial cells closest to the VEGF-A source would be assumed to stand a higher chance than their neighbours of becoming induced as tip cells, it is nevertheless difficult to conceive how a graded distribution of VEGF-A alone—an 'analogue' signal—would suffice to select singular tip cells to lead each sprout—a 'digital' response. Our loss- and gain-of function analyses together with the localization of the Dll4 signal and Notch activity suggest a model in which Dll4–Notch1 signalling between endothelial cells functions equivalent to an analogue-to-digital converter, establishing cell boundaries and the selection of either tip (low Notch signal) or stalk (high Notch signal) phenotypes within the population of VEGF-A-stimulated endothelial cells (Supplementary Fig. 11). The fundamental importance of the selection of correct numbers of endothelial tip cells, sprouts and branches is consistent with the unique genetic haploinsufficiency for the two critical ligands that regulate the process: VEGF and Dll4. VEGF inhibition is so far the only clinically validated anti-angiogenic strategy, and as such applicable to pathological angiogenesis both in tumours and ocular diseases[30]. Our present data suggest that GSIs or other modulators of Dll4–Notch1 signalling might find similar usage in the clinic.

METHODS

Animal procedures. Pharmacological Notch inhibition or activation was induced by subcutaneously injecting GSIs and Jag1 peptides, respectively, into newborn mice, followed by analysis at different time points during the first postnatal week. Retinal angiogenesis in genetic loss-of-function of Dll4 was studied in postnatally surviving $Dll4^{+/-}$ mice on an ICR background. Inducible endothelium-specific inactivation of Notch1 was obtained using intragastric injections of tamoxifen into VEcad-CreER^{T2}/R26R/Notch1$^{floxed/floxed}$ mice on a congenic C57Bl6J background. Pathological eye angiogenesis was induced by exposure to 80% oxygen during P7–12 followed by a return to normoxia and analysis at P17.

Tissue imaging. Standard protocols and commercially available antibodies were used in most instances to whole-mount-stain fixed and dissected retinas. Combined nuclear- and isolectin B4 labelling was done to allow discrimination between nuclei belonging to endothelial tip and stalk cells. Confocal laser scanning microscopy was performed using Zeiss LSM 510 Meta microscopes.

Molecular analyses. The MMV00 mouse VEGF-A ELISA kit (R&D Systems) was used to quantify VEGF-A protein concentration in tissues lysed in NP-40-based buffer. Quantitative real-time PCR was performed on tissues collected in RNA*later* (Qiagen) before and during dissection of retinas. Total retinal RNA was isolated using the RNeasy Mini Kit (Qiagen) and assays were run using TaqMan solutions and gene expression assays specific for individual mouse genes (Applied Biosystems) on an ABI Prism 7700 Sequence Detection System.

Received 16 September 2006; accepted 5 January 2007.
Published online 28 January 2007.

1. Ruhrberg, C. *et al.* Spatially restricted patterning cues provided by heparin-binding VEGF-A control blood vessel branching morphogenesis. *Genes Dev.* **16,** 2684–2698 (2002).
2. Gerhardt, H. *et al.* VEGF guides angiogenic sprouting utilizing endothelial tip-cell filopodia. *J. Cell Biol.* **161,** 1163–1177 (2003).
3. Gale, N. W. *et al.* Haploinsufficiency of delta-like 4 ligand results in embryonic lethality due to major defects in arterial and vascular development. *Proc. Natl Acad. Sci. USA* **101,** 15949–15954 (2004).
4. Krebs, L. T. *et al.* Haploinsufficienct lethality and formation of arteriovenous malformations in Notch pathway mutants. *Genes Dev.* **18,** 2469–2473 (2004).
5. Duarte, A. *et al.* Dosage-sensitive requirement for mouse Dll4 in artery development. *Genes Dev.* **18,** 2474–2478 (2004).
6. Uv, A., Cantera, R. & Samakovlis, C. *Drosophila* tracheal morphogenesis: intricate cellular solutions to basic plumbing problems. *Trends Cell Biol.* **13,** 301–309 (2003).
7. Carmeliet, P. & Tessier-Lavigne, M. Common mechanisms of nerve and blood vessel wiring. *Nature* **436,** 193–200 (2005).
8. Lu, X. *et al.* The netrin receptor UNC5B mediates guidance events controlling morphogenesis of the vascular system. *Nature* **432,** 179–186 (2004).
9. Torres-Vazquez, J. *et al.* Semaphorin-plexin signaling guides patterning of the developing vasculature. *Dev. Cell* **7,** 117–123 (2004).
10. Llimargas, M. The Notch pathway helps to pattern the tips of the *Drosophila* tracheal branches by selecting cell fates. *Development* **126,** 2355–2364 (1999).
11. Steneberg, P., Hemphala, J. & Samakovlis, C. Dpp and Notch specify the fusion cell fate in the dorsal branches of the *Drosophila* trachea. *Mech. Dev.* **87,** 153–163 (1999).
12. Ghabrial, A. S. & Krasnow, M. A. Social interactions among epithelial cells during tracheal branching morphogenesis. *Nature* **441,** 746–749 (2006).
13. Shawber, C. J. & Kitajewski, J. Notch function in the vasculature: Insights from zebrafish, mouse and man. *Bioessays* **26,** 225–234 (2004).
14. Dovey, H. F. *et al.* Functional γ-secretase inhibitors reduce β-amyloid peptide levels in brain. *J. Neurochem.* **76,** 173–181 (2001).
15. Searfoss, G. H. *et al.* Adipsin, a biomarker of gastrointestinal toxicity mediated by a functional γ-secretase inhibitor. *J. Biol. Chem.* **278,** 46107–46116 (2003).
16. van Es, J. H. *et al.* Notch/γ-secretase inhibition turns proliferative cells in intestinal crypts and adenomas into goblet cells. *Nature* **435,** 959–963 (2005).
17. Kopan, R. & Ilagan, M. X. G. γ-secretases: proteasome of the membrane? *Nature Rev. Mol. Cell Biol.* **5,** 499–504 (2004).
18. Claxton, S. & Fruttiger, M. Periodic Delta-like 4 expression in developing retinal arteries. *Gene Expr. Patterns* **5,** 123–127 (2004).
19. Shutter, J. R. *et al.* Dll4, a novel Notch ligand expressed in arterial endothelium. *Genes Dev.* **14,** 1313–1318 (2000).
20. Lawson, N. D. *et al.* Notch signaling is required for arterial-venous differentiation during embryonic vascular development. *Development* **128,** 3675–3683 (2001).
21. Domenga, V. *et al.* Notch3 is required for arterial identity and maturation of vascular smooth muscle cells. *Genes Dev.* **18,** 2730–2735 (2004).
22. Krebs, L. T. *et al.* Notch signaling is essential for vascular morphogenesis in mice. *Genes Dev.* **14,** 1343–1352 (2000).
23. Limbourg, F. P. *et al.* Essential role of endothelial Notch1 in angiogenesis. *Circulation* **111,** 1826–1832 (2005).
24. Monvoisin, A. *et al.* VE-cadherin-CreER^{T2} transgenic mouse: A model for inducible recombination in the endothelium. *Dev. Dyn.* **235,** 3413–3422 (2006).
25. Wolfer, A. *et al.* Inactivation of Notch1 in immature thymocytes does not perturb CD4 or CD8 T cell development. *Nature Immunol.* **2,** 235–241 (2001).
26. Duncan, A. W. *et al.* Integration of Notch and Wnt signaling in hematopoietic stem cell maintenance. *Nature Immunol.* **6,** 314–322 (2005).
27. Itoh, M. *et al.* Mind Bomb is a ubiquitin ligase that is essential for efficient activation of Notch signaling by Delta. *Dev. Cell* **4,** 67–82 (2003).
28. Weijzen, S. *et al.* The notch ligand Jagged-1 is able to induce maturation of monocyte-derived human dendritic cells. *J. Immunol.* **169,** 4273–4278 (2002).
29. Grunstein, J., Masbad, J. J., Hickey, R., Giordano, F. & Johnson, R. S. Isoforms of vascular endothelial growth factor act in a coordinate fashion to recruit and expand tumor vasculature. *Mol. Cell. Biol.* **20,** 7282–7291 (2000).
30. Ferrara, N. Vascular endothelial growth factor: Basic science and clinical progress. *Endocr. Rev.* **25,** 581–611 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank F. Radke for providing Notch1$^{floxed/floxed}$ mice. Support from the following foundations and granting agencies is acknowledged: Swedish Cancer Society, Association for International Cancer Research, European Union, the Novo Nordisk, Strategic Research, Söderberg, Hedlund, Wallenberg and Inga-Britt and Arne Lundberg Foundations (to C.B.); National Institutes of Health (US, NIH) and JH (USPHS National Research Service Award) (to L.I.-A.). H.G., L.-K.P. and P.L. are supported by Cancer Research UK. We acknowledge the Swegene Centre for Cellular Imaging at Gothenburg University for the use of imaging equipment, and the Light Microscopy Service and Peptide Synthesis Laboratory, London Research Institute (Cancer Research UK) for technical assistance.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.H. (mats.hellstrom@ki.se) or H.G. (holger.gerhardt@cancer.org.uk).

LETTERS

Notch signalling limits angiogenic cell behaviour in developing zebrafish arteries

Arndt F. Siekmann[1] & Nathan D. Lawson[1]

Recent evidence indicates that growing blood-vessel sprouts consist of endothelial cells with distinct cell fates and behaviours[1,2]; however, it is not clear what signals determine these sprout cell characteristics. Here we show that Notch signalling is necessary to restrict angiogenic cell behaviour to tip cells in developing segmental arteries in the zebrafish embryo. In the absence of the Notch signalling component Rbpsuh (recombining binding protein suppressor of hairless) we observed excessive sprouting of segmental arteries, whereas Notch activation suppresses angiogenesis. Through mosaic analysis we find that cells lacking Rbpsuh preferentially localize to the terminal position in developing sprouts. In contrast, cells in which Notch signalling has been activated are excluded from the tip-cell position. *In vivo* time-lapse analysis reveals that endothelial tip cells undergo a stereotypical pattern of proliferation and migration during sprouting. In the absence of Notch, nearly all sprouting endothelial cells exhibit tip-cell behaviour, leading to excessive numbers of cells within segmental arteries. Furthermore, we find that *flt4* (*fms-related tyrosine kinase 4*, also called *vegfr3*) is expressed in segmental artery tip cells and becomes ectopically expressed throughout the sprout in the absence of Notch. Loss of *flt4* can partially restore normal endothelial cell number in Rbpsuh-deficient segmental arteries. Finally, loss of the Notch ligand *dll4* (*delta-like 4*) also leads to an increased number of endothelial cells within segmental arteries. Together, these studies indicate that proper specification of cell identity, position and behaviour in a developing blood-vessel sprout is required for normal angiogenesis, and implicate the Notch signalling pathway in this process.

To investigate the role of Notch signalling during angiogenesis, we observed segmental artery formation in transgenic zebrafish embryos with fluorescent blood vessels (referred to as *Tg(fli1:egfp)^{y1}* embryos)[3]. In *Tg(fli1:egfp)^{y1}* embryos injected with a scrambled control morpholino oligonucleotide, segmental arteries emanate from the dorsal aorta and reach the dorsal-most aspect of the somite to form the dorsal longitudinal anastomotic vessel (DLAV) at 30 h post fertilization (30 hpf; Fig. 1a, b, g). Segmental arteries in *Tg(fli1:egfp)^{y1}* embryos injected with a morpholino oligonucleotide that blocks translation of Rbpsuh (a DNA-binding protein that interacts with the Notch intracellular domain and mediates transcription of Notch target genes[4]) display excessive branch points leading to unconnected endothelial tips or aberrantly fused blood vessels compared with those in control morpholino-oligonucleotide-injected siblings (Fig. 1a–d, Supplementary Table 1 and Supplementary Fig. 1). Rbpsuh-deficient embryos also display other cardiovascular defects associated with loss of Notch signalling[5], including cranial haemorrhage, loss of trunk circulation, and loss of arterial marker gene expression (Supplementary Table 2 and Supplementary Fig. 2). In contrast to loss of Notch function, activation of Notch at the 18-somite stage with an inducible Gal4/UAS system[5,6] blocks sprouting of segmental arteries (Fig. 1e, f), a phenotype similar to embryos lacking components of the Vegf signalling pathway[7–9]. Occasional sprouting cells can be seen in embryos with activated Notch (Fig. 1f, arrowhead), and we often observed cytoplasmic extensions at the position where segmental arteries would normally form (Fig. 1f, arrow), suggesting that cells with activated Notch fail to migrate in response to pro-angiogenic cues that normally promote segmental artery formation.

At 30 hpf, segmental arteries consist of three or four endothelial cells with distinct positional fates[1]: a dorsal-most T-shaped cell that contributes to the DLAV, an adjacent connector cell that lies along the length of the medial somite boundary, and a base cell that connects to the dorsal aorta (Fig. 1g). Because the Notch pathway governs cell fate decisions[10], we investigated whether these segmental artery fates were affected by Notch signalling. We transplanted cells from *Tg(fli1:egfp)^{y1}* embryos either lacking Rpbsuh or expressing an activated form of Notch into non-transgenic wild-type hosts at sphere stage and assayed the contribution of donor cells to different trunk blood-vessel types (Fig. 1g). At 30 hpf, we observed that wild-type donor *Tg(fli1:egfp)^{y1}* cells were found in all trunk blood vessels (Fig. 1h and Supplementary Table 3). In contrast, no Rbpsuh-deficient donor cells were found in the dorsal aorta at 30 hpf; instead they showed a preference for the most dorsal (DLAV) position or the posterior cardinal vein (Fig. 1i and Supplementary Table 3). Despite the requirement of Notch for artery differentiation[5,11], green fluorescent protein (GFP)-positive Rbpsuh deficient cells did initially enter the dorsal aorta at earlier time points (24 hpf; see Supplementary Fig. 3 and Supplementary Movie 1), but then migrated into the segmental artery. We also observed a contribution of GFP-positive donor cells within the dorsal aorta to haematopoietic cells, consistent with the expression of the *fli1:egfp* transgene in blood cells[3] (Supplementary Fig. 4 and Supplementary Movie 1). In contrast, transplanted cells in which Notch signalling was activated never populated the DLAV and preferentially located to the dorsal aorta or the base cell of developing sprouts (Fig. 1j and Supplementary Table 3). These cells often extended cytoplasmic protrusions into the sprout without the cell body migrating dorsally (Fig. 1j, left panel, arrows), resembling cells in embryos with global Notch activation (Fig. 1f). Although transplanted cells were also capable of contributing to non-endothelial tissues (for example, see Supplementary Fig. 3), the position and type of these cells did not correlate with donor endothelial cell position in segmental artery sprouts (data not shown). Taken together, these results suggest that Notch signalling is required cell-autonomously to determine proper endothelial cell fate within developing segmental artery sprouts.

To investigate the cellular defects caused by loss of Notch signalling, we observed segmental artery formation by confocal time-lapse analysis in live transgenic embryos bearing a nuclear localized *egfp* transgene driven by the zebrafish *fli1* promoter (*Tg(fli1:negfp)^{y7}*)[12]. In wild-type *Tg(fli1:negfp)^{y7}* embryos we observed largely stereotypical

[1]Program in Gene Function and Expression, University of Massachusetts Medical School, Lazare Research Building, 364 Plantation Street, Worcester, Massachusetts 01605, USA.

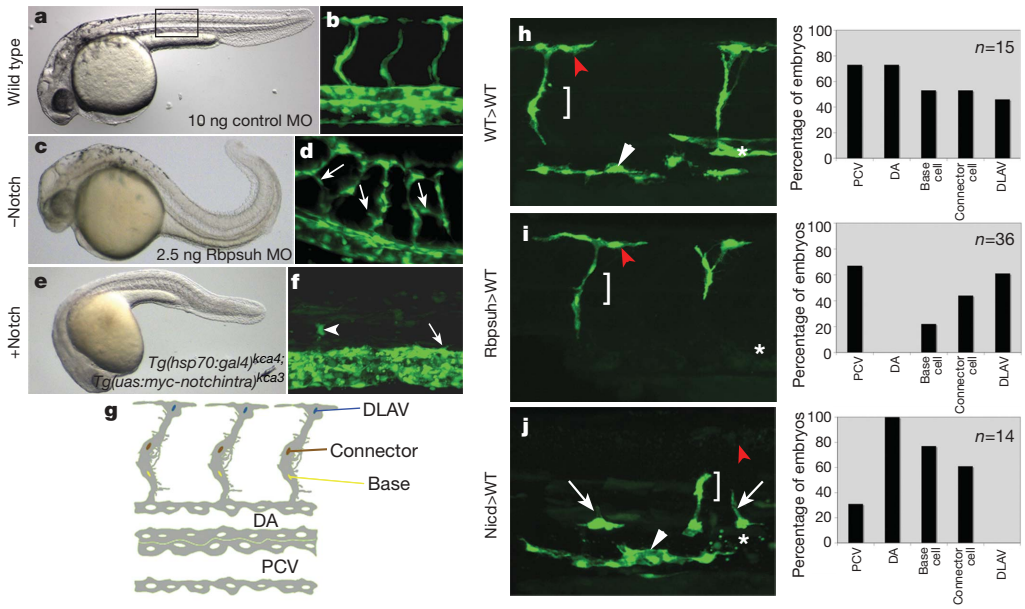


Figure 1 | Notch signalling influences positional cell fates in segmental artery sprouts. **a**–**f**, *Tg(fli:egfp)*^{y1} embryos at the 30 hpf stage. **a**, **c**, **e**, Bright-field images. **b**, **d**, **f**, Confocal images. **a**–**j**, Lateral views, dorsal is up, anterior to the left. The boxed region in **a** indicates the imaged area in **b**, **d**, **f**, **g**–**j**. **a**, **b**, General morphology (**a**) and segmental artery sprouts (**b**) in a *Tg(fli1:egfp)*^{y1} embryo injected with 10 ng control morpholino oligonucleotide (MO). **c**, **d**, Embryo injected with 2.5 ng Rbpsuh morpholino oligonucleotide. **d**, Ectopic sprouts in a Rbpsuh-deficient embryo (arrows). **e**, **f**, Notch activation after heat shock in a *Tg(hsp70:gal4)*^{kca4};*(uas:myc-notch1aintra)*^{kca3};*(fli1:egfp)*^{y1} embryo. Occasional sprouts (arrowhead) or cytoplasmic extensions of dorsal aorta cells into sprouts (arrow) can be observed. **g**, Schematic to indicate cell fates within trunk blood vessels in a zebrafish embryo at 30 hpf. **h**–**j**, Confocal images (left panels) and quantification (right panels) of donor *Tg(fli:egfp)*^{y1} cells in non-transgenic embryos. **h**, Wild-type cells contribute to all trunk vessels, including dorsal aorta (asterisk), posterior cardinal vein (white arrowhead), connector cells (bracket) and DLAV cells (red arrowhead). **i**, Endothelial cells lacking Notch signalling fail to contribute to the dorsal aorta (asterisk), and preferentially migrate into segmental artery sprouts (bracket and red arrowhead). **j**, Cells with activated Notch can be found in the dorsal aorta (asterisk), posterior cardinal vein (white arrowhead) and the connector cell (bracket), but are excluded from the DLAV (red arrowhead). DA, dorsal aorta; DLAV, dorsal longitudinal anastomotic vessel; Nicd, Notch intracellular domain; PCV, posterior cardinal vein.

behaviour of endothelial cells during segmental artery formation (Fig. 2a and Supplementary Movie 2). Cells exit the dorsal aorta and migrate between the somite boundaries to initiate sprouting at about 20 hpf (Fig. 2a, 19:49). Once cells reach the horizontal myoseptum, the segmental artery tip cell undergoes a single cell division (Fig. 2a, 20:38 and 23:22), after which one cell maintains its position and becomes a connector cell while the tip cell continues to migrate dorsally and forms the DLAV (Fig. 2a, 25:16). By 30 hpf, segmental arteries in wild-type *Tg(fli1:negfp)*^{y7} embryos consisted of three or four cells (Fig. 2a, 31:33 and see Fig. 3l). In *Tg(fli1:negfp)*^{y7} embryos injected with Rbpsuh morpholino oligonucleotide, sprout formation was initiated at a similar time point as that observed in control embryos (Fig. 2b and Supplementary Movie 3). However, more cells contributed to the initial sprout through migration (Fig. 2b, 23:06); by 24 hpf, five cells had migrated into the developing segmental artery (Fig. 2b, 24:11), exceeding the number of cells normally found in control sprouts at 30 hpf (Fig. 2a, 31:33 and Fig. 3l). Although only the tip cell undergoes a cell division in control embryos, we observed that numerous cells proliferate in Rbpsuh-deficient embryos (Fig. 2b, 24:11 through to 29:39). Furthermore, most cells within the segmental artery continued to migrate dorsally, leading to excessive cells within the DLAV at 30 hpf (compare Fig. 2a 31:33 and Fig. 2b 29:39). Thus, supernumerary endothelial cells in Rbpsuh-deficient embryos display the proliferative and migratory behaviour that is normally restricted to the leading tip cells in wild-type segmental artery sprouts. This observation suggests that Notch signalling normally

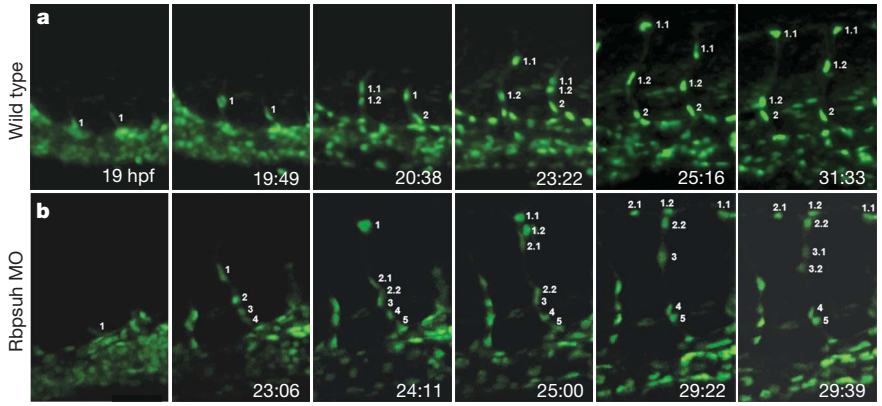


Figure 2 | Confocal time lapses of wild-type and Rbpsuh-deficient embryos. **a**, Wild-type *Tg(fli1:negfp)*^{y7} embryo from 19 hpf to 31 hpf (see Supplementary Movie 2). **b**, *Tg(fli1:negfp)*^{y7} embryo injected with 2.5 ng Rbpsuh morpholino oligonucleotide (MO) imaged from 19 hpf to 30 hpf (see Supplementary Movie 3). Cells designated with decimal points indicate progeny cells.

limits these cellular behaviours to a single tip cell in the developing sprout.

In addition to stereotypical tip-cell behaviours, we find that expression of the zebrafish orthologue of VEGF receptor-3, *flt4*, is restricted to segmental artery tip cells at 24 hpf (Fig. 3a–d, black arrowheads in panels a and c; for reference, see the 23:22 time point in Fig. 2a). Upon initial formation of the DLAV (28 hpf), *flt4* expression is no longer detectable in these blood vessels (Fig. 3e); *flt4* is also expressed in the posterior cardinal vein but not the dorsal aorta (Fig. 3d, e, blue brackets), as we and others have previously noted[5,13,14]. In Rbpsuh-deficient embryos, we observed ectopic *flt4* expression in the dorsal aorta at 24 hpf (Fig. 3f, red bracket) and expression is apparent in both tip cells (black arrowheads in Fig. 3f) and in cells sprouting from the dorsal aorta (white arrow, Fig. 3f). At 28 hpf, *flt4* expression persists in segmental arteries (Fig. 3g, white arrows) and the DLAV (Fig. 3g, black arrowheads) of Rbpsuh-deficient embryos. Because *flt4* is required for segmental artery morphogenesis in zebrafish[7] and is important for endothelial cell migration and proliferation[15], we determined whether *flt4* upregulation contributed to excessive segmental artery cell number associated with loss of Rbpsuh. For this purpose, we reduced Flt4 levels in Rbpsuh-deficient embryos and quantified endothelial cell numbers in segmental arteries. In control *Tg(fli1:negfp)^{y1}* embryos at 30 hpf, we observed that most segmental arteries contained three or four cells and occasionally up to five cells (Fig. 3h, l and Supplementary Table 4). Embryos injected with *flt4* morpholino oligonucleotide[7] alone displayed reduced cell numbers (Fig. 3j, l, Supplementary Table 4 and Supplementary Fig. 5), and these cells often failed to migrate beyond the horizontal myoseptum. In contrast, embryos injected with Rbpsuh morpholino oligonucleotide displayed segmental arteries with increased cell numbers (Fig. 3i, l, Supplementary Table 4 and Supplementary Fig. 5). In embryos co-injected with *flt4* and Rbpsuh morpholino oligonucleotides, sprout cells reached the DLAV (Fig. 3k), but many embryos displayed normal endothelial cell numbers in segmental arteries, suggesting a partial rescue of the Rbpsuh morpholino oligonucleotide phenotype (Fig. 3k, l, Supplementary Table 4 and Supplementary Fig. 5). Taken together, these results suggest that Notch activation might normally repress *flt4* to limit angiogenic cell behaviour in developing segmental artery sprouts. This possibility is consistent with our previous finding that ectopic Notch activation represses *flt4* expression in all blood vessels[5]. Notably, Notch can similarly inhibit VEGF signalling in mammalian endothelial cell lines, although in these cases VEGF receptor-2/*Kdr* is repressed by Notch activation[16,17].

Notch receptors are activated by binding to their cognate ligands on neighbouring cells[10]. We previously found that transcripts for the *notch3* receptor are expressed within the dorsal aorta in zebrafish embryos[5]. Similarly, we found that the zebrafish orthologue for the Notch ligand *dll4* is expressed in the dorsal aorta and segmental artery sprouts (Fig. 4a, b), similar to its expression in mouse embryos[18]. To

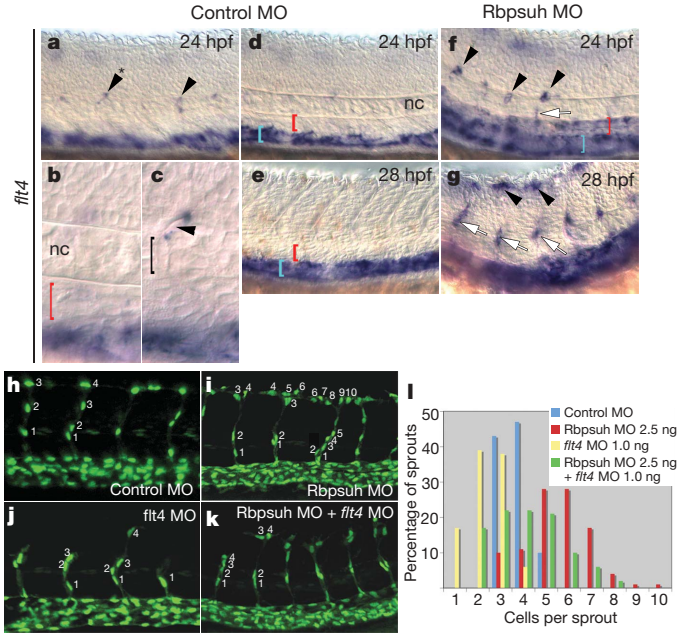


Figure 3 | Endothelial cell numbers in segmental artery sprouts and expression of *flt4* depend on Notch signalling. **a**–**g**, Whole-mount *in situ* hybridization analysis of *flt4* expression in control and Rbpsuh morpholino-oligonucleotide-injected embryos. **a**–**e**, Embryo injected with control morpholino oligonucleotide at 24 hpf. **a**, *flt4* expression in the tip cells (black arrowheads) of the developing segmental arteries (see Fig. 2a time point 23:22 for reference). Asterisk denotes cell shown in **c**. **b**, Higher magnification of dorsal aorta (red bracket) and notochord (nc) in embryo shown in **a**. **c**, Focal plane above that in **b** showing *flt4* expression in a segmental artery tip cell (black arrowhead) but not in an adjacent cell (black bracket). **d**, From the same embryo as **a** but viewed at a lower focal plane to visualize dorsal aorta (indicated by red bracket) and posterior cardinal vein (blue bracket). **e**, *flt4* expression in posterior cardinal vein (indicated by a blue bracket), but not in the dorsal aorta or segmental arteries, in an embryo injected with control morpholino oligonucleotide at 28 hpf. **f**, Embryo injected with 2.5 ng Rbpsuh morpholino oligonucleotide at 24 hpf. *flt4* expression is present in segmental artery cells (black arrowheads), dorsal aorta (red bracket), posterior cardinal vein (blue bracket) and in cells emanating from the dorsal aorta (white arrow). **g**, Embryo injected with 2.5 ng Rbpsuh morpholino oligonucleotide at 28 hpf. *flt4* expression is present in segmental artery connector cells (white arrows) and in the DLAV (black arrowheads). **h**–**k**, Confocal images of *Tg(fli1:negfp)^{y7}* embryos at 30 hpf. **h**, Control morpholino-oligonucleotide-injected *Tg(fli:negfp)^{y7}* embryo. **i**, Segmental arteries in an embryo injected with 2.5 ng Rbpsuh morpholino oligonucleotide contain excessive endothelial cells. **j**, Injection of *flt4* morpholino oligonucleotide reduces endothelial cell number. **k**, Embryo co-injected with morpholino oligonucleotides against Rbpsuh and *flt4*, showing formation of DLAV and normal segmental artery cell number. **l**, Quantification of endothelial cell number in segmental artery sprouts. **a**–**k**, Lateral views, anterior to the left, dorsal is up.

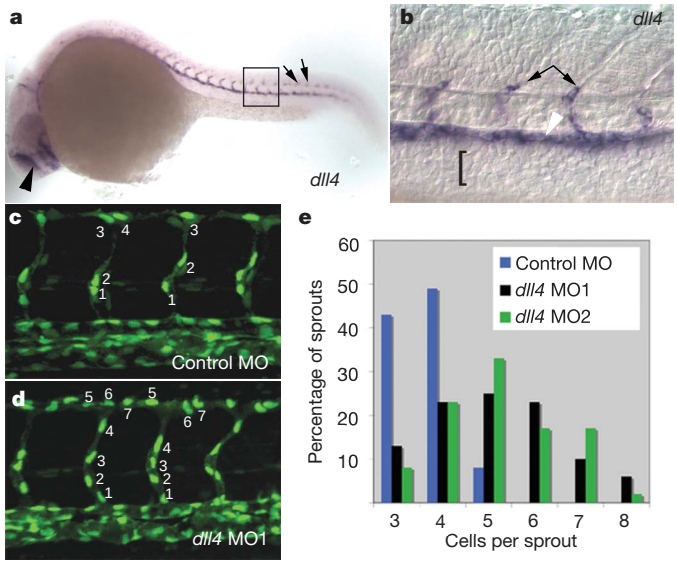


Figure 4 | Loss of *dll4* phenocopies Rbpsuh loss of function. **a**, *In situ* hybridization at 24 hpf showing *dll4* expression in the dorsal aorta and segmental arteries. Expression in the telencephalon (arrowhead) and neurons (arrows) is apparent; the box indicates view in **b**. **b**, Close up of the embryo in **a**. *dll4* is restricted to the dorsal aorta (white arrowhead) and segmental artery sprouts (black arrows; bracket indicates posterior cardinal vein). **c**, **d**, Confocal images of *Tg(fli:negfp)^{y7}* embryos at 30 hpf. **c**, Control morpholino-oligonucleotide-injected embryos. **d**, Embryos injected with *dll4* MO1 show increased segmental artery cell number. **e**, Quantification of segmental artery cell number in *dll4* morpholino-oligonucleotide-injected embryos. (See Supplementary Fig. 6 for explanation of MO1 and MO2.)

determine whether loss of *dll4* phenocopied the segmental artery defects associated with loss of Rbpsuh described above, we designed morpholino oligonucleotides that prevented splicing of *dll4* pre-messenger RNA (see Supplementary Fig. 6). *Tg(fli1:negfp)^{y7}* embryos lacking *dll4* showed an increase in endothelial cell number within segmental arteries compared with embryos injected with control morpholino oligonucleotide (Fig. 4c–e, Supplementary Table 5 and Supplementary Fig. 7). Despite this defect, vascular morphology and cardiovascular function were grossly normal until day 2 of development, although *dll4*-deficient embryos subsequently displayed ectopic segmental vessel branching similar to Rbpsuh-deficient embryos (Supplementary Fig. 1). The finding that loss of *dll4*, an artery-specific Notch ligand, phenocopies the cellular defects observed in embryos globally deficient for Rbpsuh further supports an endothelial cell-autonomous role for Notch signalling in limiting angiogenesis. Notably, loss of *dll4* causes milder vascular morphology defects and has no apparent effect on artery differentiation when compared with Rbpsuh-deficient embryos, suggesting the role of additional Notch ligands. This would be consistent with the widespread expression of numerous Notch ligands and receptors in endothelial cells[19,20], and indicates that distinct Notch receptor–ligand pairs may mediate different endothelial cell fate decisions or cell behaviours. Furthermore, these results suggest that the role of Notch in determination of arteriovenous cell fates[5,11] may be distinct from its role in limiting angiogenic blood vessel growth, although we cannot rule out the possibility that these two processes are more intimately connected.

Angiogenesis requires the coordination of distinct cellular behaviours in adjacent endothelial cells in order to form a patent blood vessel. Whereas the sprout tip cell undergoes migration and proliferation, the adjacent trailing cells must maintain their positions in order to form a lumen and connect to the existing circulatory network. These distinct behaviours are apparent despite the probable exposure of multiple sprout cells to pro-angiogenic factors, such as Vegf. Our results suggest that the Notch signalling pathway has an important function in controlling normal angiogenesis by limiting angiogenic cell behaviours to the endothelial tip cell, possibly through the negative regulation of Vegf receptor signalling. Of note, our previous work has demonstrated that the Notch and Vegf signalling pathways act in a common pathway to define arterial endothelial cell fate[13]. Here, we find that Notch negatively regulates Vegf-dependent angiogenesis. These observations suggest that Notch may act as an important switch to modulate Vegf signalling in different settings within the developing vasculature.

METHODS

Zebrafish were maintained according to standard methods[21]. The establishment and characterization of transgenic lines used in this study have been described elsewhere[3,6,12,22]. Mosaic analysis was performed using cell transplantation as described[21]. Endothelial cell contribution of transplanted cells was assayed by virtue of EGFP expression in chimaeric embryos. A zebrafish *dll4* orthologue (GenBank accession number CAI11901) was identified through database searches, and a corresponding antisense riboprobe was generated using previous methods[7]. The Rbpsuh and *flt4* morpholino oligonucleotides have been described previously[23] and sequences for *dll4* and scrambled morpholino oligonucleotides are available in Supplementary Methods. Whole-mount *in situ* hybridization was carried out as described[5]. Time-lapse analyses were performed as described elsewhere[3]. Confocal stacks and movies were assembled using Imaris Software (Bitplane) and movies were labelled using Flash Player (Macromedia). Confocal micro-angiography was performed as described[24]. To quantify segmental artery cell numbers, we counted cell nuclei using the *Tg(fli1:negfp)^{y7}* line[12]. For the graphs displayed in Supplementary Tables 3 and 4, we averaged total cell counts from each embryo and calculated standard deviation and *P*-values using a two-tailed *t*-test. Detailed experimental methods are available in Supplementary Methods.

Received 16 October 2006; accepted 8 January 2007.
Published online 28 January 2007.

1. Childs, S., Chen, J. N., Garrity, D. M. & Fishman, M. C. Patterning of angiogenesis in the zebrafish embryo. *Development* **129,** 973–982 (2002).
2. Gerhardt, H. *et al.* VEGF guides angiogenic sprouting utilizing endothelial tip cell filopodia. *J. Cell Biol.* **161,** 1163–1177 (2003).
3. Lawson, N. D. & Weinstein, B. M. *In vivo* imaging of embryonic vascular development using transgenic zebrafish. *Dev. Biol.* **248,** 307–318 (2002).
4. Lecourtois, M. & Schweisguth, F. The neurogenic suppressor of hairless DNA-binding protein mediates the transcriptional activation of the enhancer of split complex genes triggered by Notch signaling. *Genes Dev.* **9,** 2598–2608 (1995).
5. Lawson, N. D. *et al.* Notch signaling is required for arterial-venous differentiation during embryonic vascular development. *Development* **128,** 3675–3683 (2001).
6. Scheer, N. & Campos-Ortega, J. A. Use of the Gal4-UAS technique for targeted gene expression in the zebrafish. *Mech. Dev.* **80,** 153–158 (1999).
7. Covassin, L. D., Villefranc, J. A., Kacergis, M. C., Weinstein, B. M. & Lawson, N. D. Distinct genetic interactions between multiple Vegf receptors are required for development of different blood vessel types in zebrafish. *Proc. Natl Acad. Sci. USA* **103,** 6554–6559 (2006).
8. Lawson, N. D., Mugford, J. W., Diamond, B. A. & Weinstein, B. M. phospholipase Cγ1 is required downstream of vascular endothelial growth factor during arterial development. *Genes Dev.* **17,** 1346–1351 (2003).
9. Nasevicius, A., Larson, J. & Ekker, S. C. Distinct requirements for zebrafish angiogenesis revealed by a VEGF-A morphant. *Yeast* **17,** 294–301 (2000).
10. Artavanis-Tsakonas, S., Rand, M. D. & Lake, R. J. Notch signaling: cell fate control and signal integration in development. *Science* **284,** 770–776 (1999).
11. Krebs, L. T. *et al.* Haploinsufficient lethality and formation of arteriovenous malformations in Notch pathway mutants. *Genes Dev.* **18,** 2469–2473 (2004).
12. Roman, B. L. *et al.* Disruption of acvrl1 increases endothelial cell number in zebrafish cranial vessels. *Development* **129,** 3009–3019 (2002).
13. Lawson, N. D., Vogel, A. M. & Weinstein, B. M. sonic hedgehog and vascular endothelial growth factor act upstream of the Notch pathway during arterial endothelial differentiation. *Dev. Cell* **3,** 127–136 (2002).
14. Thompson, M. A. *et al.* The *cloche* and *spadetail* genes differentially affect hematopoiesis and vasculogenesis. *Dev. Biol.* **197,** 248–269 (1998).
15. Salameh, A., Galvagni, F., Bardelli, M., Bussolino, F. & Oliviero, S. Direct recruitment of CRK and GRB2 to VEGFR-3 induces proliferation, migration, and survival of endothelial cells through the activation of ERK, AKT, and JNK pathways. *Blood* **106,** 3423–3431 (2005).
16. Taylor, K. L., Henderson, A. M. & Hughes, C. C. Notch activation during endothelial cell network formation *in vitro* targets the basic HLH transcription factor HESR-1 and downregulates VEGFR-2/KDR expression. *Microvasc. Res.* **64,** 372–383 (2002).
17. Williams, C. K., Li, J. L., Murga, M., Harris, A. L. & Tosato, G. Up-regulation of the Notch ligand Delta-like 4 inhibits VEGF-induced endothelial cell function. *Blood* **107,** 931–939 (2006).
18. Shutter, J. R. *et al.* Dll4, a novel Notch ligand expressed in arterial endothelium. *Genes Dev.* **14,** 1313–1318 (2000).
19. Smithers, L., Haddon, C., Jiang, Y. & Lewis, J. Sequence and embryonic expression of *deltaC* in the zebrafish. *Mech. Dev.* **90,** 119–123 (2000).
20. Villa, N. *et al.* Vascular expression of Notch pathway receptors and ligands is restricted to arterial vessels. *Mech. Dev.* **108,** 161–164 (2001).
21. Westerfield, M. *The Zebrafish Book* (Univ. Oregon Press, Eugene, Oregon, 1993).
22. Scheer, N., Groth, A., Hans, S. & Campos-Ortega, J. A. An instructive function for Notch in promoting gliogenesis in the zebrafish retina. *Development* **128,** 1099–1107 (2001).
23. Sieger, D., Tautz, D. & Gajewski, M. The role of Suppressor of Hairless in Notch mediated signalling during zebrafish somitogenesis. *Mech. Dev.* **120,** 1083–1094 (2003).
24. Weinstein, B. M., Stemple, D. L., Driever, W. & Fishman, M. C. Gridlock, a localized heritable vascular patterning defect in the zebrafish. *Nature Med.* **1,** 1143–1147 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank C. Sagerström and B. Weinstein for critical reading of the manuscript. We thank J. Polli and M. Kacergis for fish care and maintenance. We would also like to thank members of the Lawson laboratory for helpful discussions, and C. Lange for the Rbpsuh *in situ* analysis. We are grateful to J. Leslie and J. Lewis for sharing data before publication. This work was supported in part by a grant awarded to N.D.L. from the National Heart Lung and Blood Institute, NIH.

Author Contributions A.F.S. and N.D.L. designed and carried out the experiments, analysed the data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.D.L (nathan.lawson@umassmed.edu).

LETTERS

Drosophila TCTP is essential for growth and proliferation through regulation of dRheb GTPase

Ya-Chieh Hsu[1], Joshua J. Chern[2], Yi Cai[4], Mingyao Liu[4] & Kwang-Wook Choi[1,2,3]

Cellular growth and proliferation are coordinated during organogenesis. Misregulation of these processes leads to pathological conditions such as cancer. Tuberous sclerosis (TSC) is a benign tumour syndrome caused by mutations in either *TSC1* or *TSC2* tumour suppressor genes. Studies in *Drosophila* and other organisms have identified TSC signalling as a conserved pathway for growth control. Activation of the TSC pathway is mediated by Rheb (Ras homologue enriched in brain), a Ras superfamily GTPase[1,2]. Rheb is a direct target of TSC2 and is negatively regulated by its GTPase-activating protein activity[3–5]. However, molecules required for positive regulation of Rheb have not been identified. Here we show that a conserved protein, translationally controlled tumour protein (TCTP), is an essential new component of the TSC–Rheb pathway. Reducing *Drosophila* TCTP (dTCTP) levels reduces cell size, cell number and organ size, which mimics *Drosophila Rheb* (*dRheb*) mutant phenotypes. *dTCTP* is genetically epistatic to *Tsc1* and *dRheb*, but acts upstream of *dS6k*, a downstream target of *dRheb*. dTCTP directly associates with dRheb and displays guanine nucleotide exchange activity with it *in vivo* and *in vitro*. Human TCTP (hTCTP) shows similar biochemical properties compared to dTCTP and can rescue *dTCTP* mutant phenotypes, suggesting that the function of TCTP in the TSC pathway is evolutionarily conserved. Our studies identify TCTP as a direct regulator of *Rheb* and a potential therapeutic target for TSC disease.

TCTP is a highly conserved protein (Supplementary Fig. 1) upregulated in various tumours. Despite studies on the biochemical and structural properties of TCTP[6–9], the physiological significance of these findings has not been determined. Thus, we aimed to study the function of TCTP *in vivo* using *Drosophila* as a model organism.

We first knocked down dTCTP expression in developing flies by targeted expression of double-stranded RNA (dsRNA) for RNA interference (RNAi)[10] using the GAL4/UAS system[11]. Expression of *dTCTP* RNAi depleted endogenous dTCTP to nearly undetectable levels by different GAL4 drivers (Fig. 1a, b). Tissue-specific expression of *dTCTP* RNAi reduced the size of the eye, wing, notum, or a specific region in the wing pouch, corresponding to the expression domains of various *GAL4* lines (Fig. 1c–e; see also Supplementary Fig. 2a–d). The size reduction was caused by a decrease in both cell size and cell number (Fig. 1f; see also Supplementary Fig. 2e–g), a typical phenotype for mutations in the insulin or TSC pathways. Ubiquitous expression of *UAS-dTCTP* RNAi by *actin-GAL4* (*act>dTCTPi*) caused lethality around the third instar larval stage. A portion of these larvae was able to survive to the pupal stage with reduced body size (data not shown), consistent with the organ size reduction. The lethality and phenotypes caused by *dTCTP* RNAi were rescued by co-expression of a *dTCTP* complementary DNA, indicating that these defects were due to a reduction of dTCTP levels. We therefore concluded that *dTCTP* RNAi suppresses organ growth by affecting both cell size and number.

We further investigated the phenotypes of *dTCTP* loss-of-function mutants because RNAi may represent a hypomorphic situation. $dTCTP^{Ey09182}$ is a hypomorphic allele of *dTCTP* resulting from a P-element insertion in its 5′ untranslated region. Rare homozygous $dTCTP^{Ey09182}$ flies that escaped from lethality showed smaller body size compared with their heterozygous siblings (Fig. 2b and Supplementary Information). To create null alleles, we generated imprecise excisions from $dTCTP^{Ey09182}$ (see Methods). One imprecise

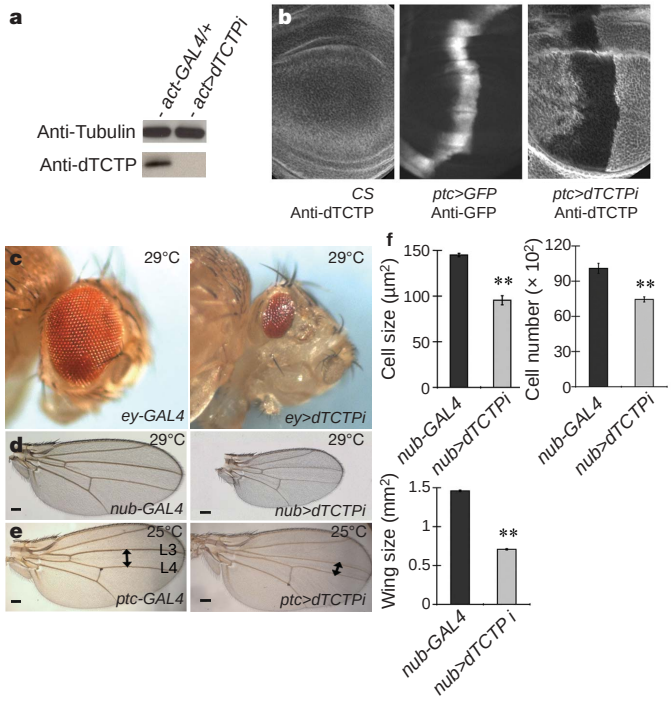


Figure 1 | *dTCTP* RNAi affects cell size, cell number and organ size. a, No dTCTP is detected in larval extract after *dTCTP* RNAi expression by *act-GAL4* at 29 °C. **b**, dTCTP is ubiquitously expressed in the wild-type (*CS*) wing disc, but is depleted by *ptc>dTCTPi* in the *ptc*-expressing region marked by GFP. **c**, Expression of *dTCTP* RNAi in eye disc (*ey>dTCTPi*) reduces the eye size. **d**, Expression of *dTCTP* RNAi in the entire wing pouch (*nub>dTCTPi*) reduces the wing size. **e**, Expression of *dTCTP* RNAi between L3 and L4 veins by *ptc-GAL4* reduces the distance between these two veins (double arrows). Scale bars in **d**, **e**, 100 μm. **f**, Quantification of defects in *nub>dTCTPi* wings ($n = 10$ for each genotype, error bars indicate s.d.; double asterisk, $P < 0.0001$). The small wing phenotype is due to smaller cell size (34% decrease) and a decreased cell number (26% decrease).

[1]Program in Developmental Biology, [2]Department of Molecular and Cellular Biology, and [3]Department of Ophthalmology, Baylor College of Medicine, and [4]Institute of Biosciences and Technology, Texas A&M University System Health Science Center, Houston, Texas 77030, USA.

excision line, *dTCTP*^{h59}, showed a 1.1-kilobase deletion downstream of the insertion site that removes the entire *dTCTP* coding sequence (Fig. 2a). Western blot analysis showed no detectable dTCTP protein in *dTCTP*^{h59} early first instar larvae (Fig. 2c). Both *dTCTP*^{h59} homozygotes and *dTCTP*^{h59} heterozygotes for a deficiency chromosome uncovering the *dTCTP* region (*dTCTP*^{h59}/*Df(3R)M-Kx1*) showed 100% lethality at the late first instar stage, indicating that this allele is a genetic null. Expression of dTCTP cDNA or a genomic DNA construct was able to rescue *dTCTP*^{h59} mutants, indicating that the lethality is due to deletion of the *dTCTP* gene (see Supplementary Information).

To examine the phenotypes of *dTCTP* null mutant cells, we generated *dTCTP* mutant clones using mitotic recombination.

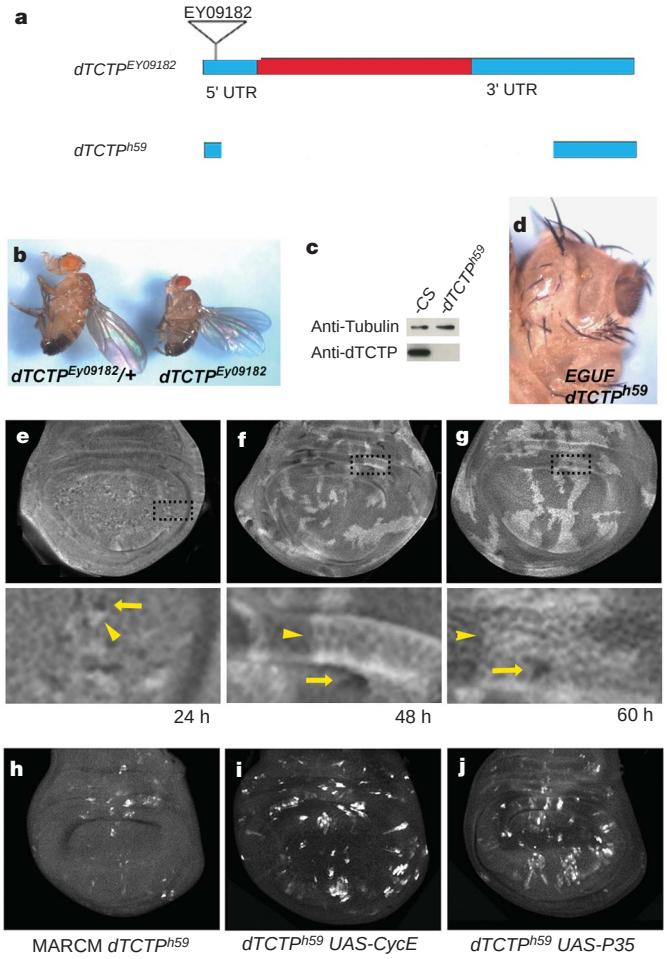


Figure 2 | *dTCTP* mutant phenotypes in cell proliferation and survival. **a**, Schematics of two *dTCTP* mutant alleles. **b**, An escaper of *dTCTP*^{EY09182} (right) and its heterozygous sibling (left). **c**, Western blot analysis detects no dTCTP protein in early first instar *dTCTP*^{h59} larvae. **d**, An eye composed of mainly *dTCTP*^{h59} cells (genotype: *w*; *ey-GAL4*, *UAS-FLP*/+; *FRT82 pGMR-hid cl3R*/*FRT82 dTCTP*^{h59}). **e**–**g**, The size of *dTCTP*^{h59} mutant clones. Discs are stained by dTCTP antibody. Clones are generated at the late second instar (**e**), the late first instar (**f**) and the mid first instar (**g**), and dissected at the late third instar stage. Bottom panels show an enlarged view of the boxed area in the top panels. Arrows and arrowheads mark *dTCTP*^{h59} homozygous mutant clones and the associated twin spots, respectively. **h**–**j**, Expression of *CycE* or *P35* within GFP$^+$ *dTCTP*^{h59} clones using the MARCM technique. The larvae are raised at 25 °C and the discs are dissected 72 h after heat shock. **h**, Most of the *dTCTP*^{h59} clones are eliminated (genotype: *hsflp, tub-GAL4, UAS-GFP*^{nls}; *FRT82 tub-GAL80/FRT82 dTCTP*^{h59}). **i**, Expression of *CycE* in *dTCTP*^{h59} clones (genotype: *hsflp, tub-GAL4, UAS-GFP*^{nls}; *UAS-CycE*/+; *FRT82 tub-GAL80/FRT82 dTCTP*^{h59}). **j**, Expression of *P35* in *dTCTP*^{h59} clones (genotype: *hsflp, tub-GAL4, UAS-GFP*^{nls}; *UAS-P35*/+; *FRT82 tub-GAL80/FRT82 dTCTP*^{h59}).

dTCTP^{h59} mutant clones showed growth disadvantage compared to their wild-type twin spots. The sizes of *dTCTP*^{h59} clones were similar to those of the twin spots 24 h after heat shock (Fig. 2e). However, the sizes of the twin spots were much larger than *dTCTP*^{h59} clones 48 h after heat shock (Fig. 2f), and most *dTCTP*^{h59} clones were eliminated by 60 h after heat shock (Fig. 2g). Similarly, using the EGUF/Hid technique[12] to remove most wild-type cells in *dTCTP*^{h59} mosaic eyes resulted in either a no-eye or small-eye phenotype (Fig. 2d). Therefore, the null phenotypes were qualitatively comparable to the *dTCTP* RNAi phenotype, but more severe.

The reduction in cell number caused by *dTCTP* RNAi and the behaviour of *dTCTP* mutant cells may result from a proliferation defect or abnormal cell death. We tested these possibilities using the MARCM (mosaic analysis with a repressible cell marker) technique[13]. Similar to clones generated by traditional mitotic recombination, numerous small *dTCTP*^{h59} green-fluorescent-protein-expressing (GFP$^+$) clones were observed at 24 h after heat shock (data not shown). By 72 h after heat shock, very few GFP$^+$ clones remained on the discs (Fig. 2h). In contrast, Cyclin E (CycE) overexpression, via the MARCM technique, within *dTCTP*^{h59} clones resulted in four times more *dTCTP*^{h59} cells at 72 h after heat shock (comparing Fig. 2i to h). Similarly, CycE overexpression also suppressed the *dTCTP* RNAi phenotypes (Supplementary Fig. 3). We next tested whether the *dTCTP*^{h59} phenotypes can be attributed to abnormal cell death. Expression of the P35 cell death inhibitor also significantly suppressed the *dTCTP*^{h59} phenotypes, leading to the presence of four times more GFP$^+$ cells at 72 h after heat shock (comparing Fig. 2j to h). These data indicate that loss of *dTCTP* causes defects in cell proliferation and triggers cell death.

Insulin and TSC signalling are two parallel but interacting pathways for growth control. Inactivation of positive regulatory components, such as *Insulin receptor* (*InR*), *dRheb* and *Tor*, leads to a decrease in organ size by affecting cell size and cell number[1,2,14–17]. In contrast, overexpression of *InR* and *dRheb*, as well as inactivation of negative regulatory components such as *Tsc1* (refs 18, 19), *Tsc2* (ref. 20) and *Pten* (refs 21–23), causes tissue overgrowth. Given the similar phenotypes between *dTCTP* mutants and mutants in the insulin/TSC pathways, we performed genetic epistasis experiments to test whether *dTCTP* has a role in these two pathways. Overexpression of *InR* by *patched* (*ptc*)*-GAL4* caused weak but consistent expansion of the distance between L3 and L4 wing veins (compare Fig. 3a, c and k). In contrast, co-expression of *InR* and *dTCTP* RNAi reduced the L3–L4 distance, resembling the *dTCTP* RNAi phenotype (compare Fig. 3b, d and k). Therefore, *dTCTP* is epistatic to *InR*.

Next, we examined the relationship between *Tsc1* and *dTCTP*. Mutations in *Tsc1* or *Tsc2* cause similar phenotypes because they function as a complex. Mosaic eyes and heads consisting primarily of *Tsc1* mutant cells were much larger than wild type (compare Fig. 3e and f). Expression of *dTCTP* RNAi in *Tsc1* mutant cells suppressed this overgrowth both in the eye and head (Fig. 3g). Furthermore, when the eye was composed of *Tsc1* and *dTCTP* double mutant cells, the flies displayed a small or no-eye phenotype indistinguishable from the *dTCTP* single mutant phenotype (compare Figs 3h and 2d), suggesting that dTCTP acts downstream or in parallel to Tsc1.

We next tested the relationship between *dTCTP* and *dRheb*. Ectopic expression of *dRheb* using *ptc–GAL4* resulted in a 15% increase in the L3–L4 distance compared with the *ptc>GFP* control (Fig. 3i, k). However, co-expression of *dTCTP* RNAi and *dRheb* in the *ptc* expression region showed the *dTCTP* RNAi phenotype (Fig. 3j, k), suggesting that *dTCTP* is epistatic to *dRheb*.

Finally, we tested whether activation of dS6k, a downstream effector of the insulin/TSC pathway, is dependent on *dTCTP*. The level of dS6k activation was measured using a phospho-specific antibody (dS6k p-Thr 398). Extracts from *act>dTCTPi* larvae showed a significantly lower amount of activated dS6k compared with the controls (Fig. 3l), indicating that *dTCTP* is required for dS6k activation. Consistent with this, the *eyeless* (*ey*)*>dTCTPi* phenotype was dominantly enhanced by heterozygosity for a null mutation of *dS6k* (*dS6k*$^{l-1}$) (Fig. 3m, n).

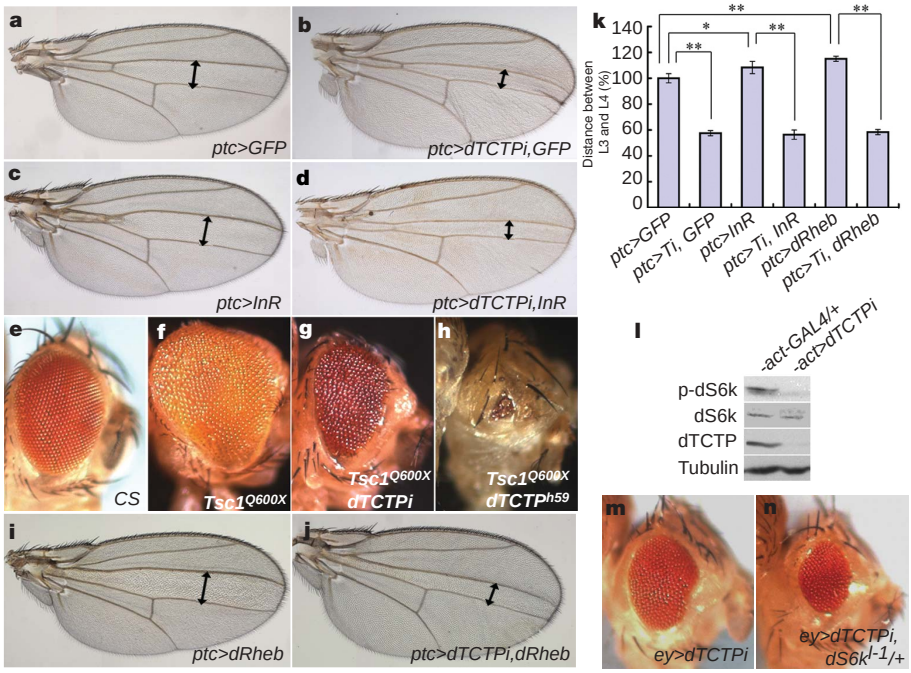


Figure 3 | Epistatic analysis between dTCTP and components of the insulin signalling and TSC pathways. Double arrows in the wings mark the L3–L4 region. **a**, A wing from a *ptc>GFP* fly as a wild-type control. **b**, *ptc>dTCTPi,GFP*. **c**, *ptc>InR*. **d**, *ptc>dTCTPi,InR*. **e**, A wild-type eye. **f**, An eye composed of only *Tsc1* null mutant cells (genotype: *w*; *ey-GAL4*, *UAS-FLP*/+; *FRT82 pGMR-hid cl3R*/*FRT82 Tsc1*^{Q600X}). **g**, Expression of *dTCTP* RNAi in the *Tsc1* mutant cells (genotype: *w*; *ey-GAL4*, *UAS-FLP*/*UAS-dTCTPi*; *FRT82 pGMR-hid cl3R*/*FRT82 Tsc1*^{Q600X}). **h**, An eye composed of *Tsc1* and *dTCTP* double mutant cells (genotype: *w*; *ey-GAL4*, *UAS-FLP*/+; *FRT82 pGMR-hid cl3R*/*FRT82 dTCTP*^{h59}, *Tsc1*^{Q600X}). **i**, *ptc>dRheb*. **j**, *ptc>dTCTPi,dRheb*. **k**, Quantification for the L3–L4 distance compared with the *ptc>GFP* control ($n = 8$ for each genotype, error bars indicate s.d.; asterisk, $P = 0.0023$; double asterisk, $P < 0.0001$). **l**, The phosphorylated dS6k level is reduced in *act>dTCTPi* flies at 18 °C. **m**, **n**, At 22 °C, *ey>dTCTPi* flies show a mild eye reduction (**m**). Removing a copy of *dS6k* enhances the phenotype (**n**).

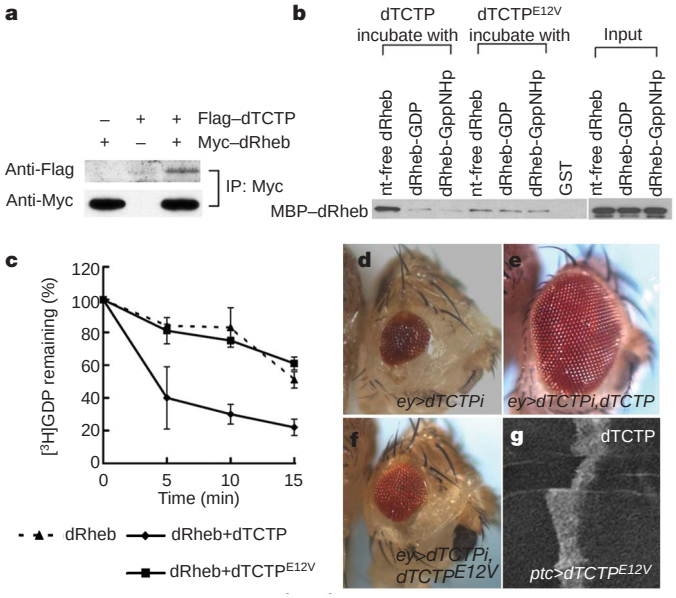


Figure 4 | dTCTP has GEF-like activity for dRheb. **a**, dTCTP and dRheb co-immunoprecipitate in 293T cells. **b**, GST–dTCTP interacts most strongly with nucleotide-free MBP–dRheb. GST– dTCTP^{E12V} can pull-down dRheb with a weaker affinity. **c**, *In vitro* GDP release assay using [^{3}H]GDP-loaded MBP–dRheb. dTCTP accelerates the GDP release from dRheb, whereas dTCTP^{E12V} shows no effect (error bars indicate s.d.; GST–dTCTP, GST–dTCTP^{E12V} or MBP–dRheb were used at 1 μM). **d**, *ey>dTCTPi* flies at 29 °C. **e**, Expression of *UAS-dTCTP* together with *dTCTP* RNAi rescues the growth defect. **f**, Expression of *UAS-dTCTP*^{E12V} fails to rescue the *dTCTP* RNAi phenotype, despite its high level of expression (**g**).

Removing a copy of *dS6k* caused an approximately 20% further reduction of eye size in *ey>dTCTPi* animals. Taken together, these data support a model wherein *dTCTP* functions either downstream or in parallel to *Tsc* and *dRheb*, but upstream of *dS6k*.

Epistatic analysis suggests that dTCTP may be a new component in the TSC pathway. Because TCTP structurally resembles a small GTPase regulator, Mss4 (ref. 9), we proposed that dTCTP might directly associate with dRheb and positively regulate its activity. To test this, we first performed co-immunoprecipitation experiments. Flag-tagged dTCTP immunoprecipitated together with Myc-tagged dRheb in 293T cell extracts (Fig. 4a), suggesting that dTCTP and dRheb form a complex *in vivo*. Furthermore, *in vitro* pull-down assays demonstrated direct binding of glutathione *S*-transferase (GST)–dTCTP to maltose binding protein (MBP)–dRheb. Notably, dTCTP showed preferential binding to nucleotide-free dRheb (Fig. 4b), a property shared among guanine nucleotide exchange factors (GEFs). To test whether dTCTP has GEF activity for dRheb, we carried out *in vitro* GDP release experiments. MBP–dRheb alone showed weak intrinsic GDP dissociation. In contrast, addition of GST–dTCTP stimulated the GTP/GDP exchange on dRheb rapidly even when low amounts of dTCTP were used (Fig. 4c; see also Supplementary Fig. 4a). The GEF-like activity is specific, as dTCTP did not accelerate the exchange reaction on dRas1, the closest GTPase to dRheb at sequence level (data not shown). Moreover, a glutamic acid to valine mutation in the putative GTPase binding groove[9] of dTCTP (dTCTP^{E12V}) abolished this GEF activity, even at a high concentration (Fig. 4c; see also Supplementary Fig. 4a). Because dTCTP^{E12V} still retained binding activity to dRheb (Fig. 4b), this residue seems to have a critical function in catalytic reactions rather than binding between the two proteins. To determine whether E12 is critical for the function of dTCTP *in vivo*, we performed genetic rescue experiments. Whereas wild-type *dTCTP* was able to rescue fully the *dTCTP* RNAi phenotype (Fig. 4d, e), mutant *dTCTP*^{E12V} failed to rescue the RNAi defects (Fig. 4f and data not shown) even

though the mutant protein was expressed at a high level (Fig. 4g). Therefore, the conserved E12 residue of dTCTP is essential for its normal function in development.

To test whether this *in vitro* GEF activity has a physiological relevance, we determined the *in vivo* level of GTP bound to dRheb in *Drosophila* S2 cells. S2 cells treated with dTCTP dsRNA or a control EGFP dsRNA were transfected with Myc-tagged dRheb. dTCTP-dsRNA-treated cells consistently showed a lower percentage of GTP-bound dRheb (Supplementary Fig. 4b), suggesting that dTCTP is required for dRheb activation *in vivo*. These GEF-like properties displayed by dTCTP towards dRheb are particularly intriguing, as no Rheb GEFs have been reported. Further kinetic and structural analysis will help to elucidate whether dTCTP is a bona fide GEF enzyme for dRheb.

Human TCTP (hTCTP) and dTCTP are roughly 50% identical in their protein sequence. We found that *dTCTP* RNAi and mutant phenotypes can be rescued by expression of *hTCTP* (Supplementary Fig. 5a–d). Furthermore, hTCTP interacted most strongly with the nucleotide-free hRheb and stimulated the GDP/GTP exchange of hRheb *in vitro* (Supplementary Fig. 5e, f). These data suggest that the function of TCTP in the TSC pathway is likely to be conserved throughout evolution.

dTCTP controls cell growth and proliferation by positively regulating dRheb activity. Our data suggest that dTCTP may function as a GEF or a related regulatory factor to activate dRheb. Given the strong epistatic effect of dTCTP to dRheb, it is also possible that dTCTP may have additional roles in the TSC pathway, acting downstream of dRheb but upstream of S6k (Supplementary Fig. S6).

TCTP has been implicated in a wide range of cancers. Nevertheless, we have not observed overgrowth phenotypes as a result of dTCTP overexpression (data not shown), suggesting that *dTCTP* is not sufficient to induce growth. Notably, reduction of TCTP levels is sufficient for suppression of malignancy in tumour reversion models[24,25]. Our study provides a possible explanation for this phenomenon. It will be intriguing to learn whether lowering the insulin/TSC signalling output can be a general mechanism for tumour reversion.

METHODS

Generation of the *dTCTP* null mutant. *dTCTP* deletion alleles were generated by imprecise mobilization of the $dTCTP^{Ey09182}$ insert. Potential excisions were identified by the loss of w^+ markers and were tested for complementation with *Df(3R)M-Kx1*, a deficiency line uncovering the *dTCTP* locus. Genomic DNA from lines that failed to complement the deficiency was used as polymerase chain reaction templates. Primers flanking the *Ey09182* insertion at position −33 and +967 relative to the insertion site were used in the reaction. A ~350-base-pair (bp) product was produced from the line $dTCTP^{h59}$. Sequencing results confirmed the left/right deletion break points at the 34- and 695-bp position of the *dTCTP* transcript, indicating a deletion of the entire *dTCTP* coding sequence. Other *Drosophila* strains and genetic crosses are detailed in Supplementary Information.

***In vitro* GST pull-down assays.** Removal of endogenous nucleotides and *in vitro* pull-down assays were performed similarly as described[26].

***In vitro* GDP release assays.** The guanine nucleotide exchange assay was performed essentially as described[27].

Measurements of the dRheb activation state *in vivo*. The percentage of activated (that is, GTP-bound) dRheb was determined based on the established methods for measuring Ras, Rap and Rheb GTPases[28–30].

Detailed information about methods used for molecular biology, immunohistochemistry, detection of phospho-dS6k levels, S2 cell RNAi, cell size and cell number measurements, and statistical analysis can be found in Supplementary Information.

Received 9 November; accepted 5 December 2006.

1. Stocker, H. *et al.* Rheb is an essential regulator of S6K in controlling cell growth in *Drosophila*. *Nature Cell Biol.* **5,** 559–565 (2003).
2. Saucedo, L. J. *et al.* Rheb promotes cell growth as a component of the insulin/TOR signalling network. *Nature Cell Biol.* **5,** 566–571 (2003).
3. Zhang, Y. *et al.* Rheb is a direct target of the tuberous sclerosis tumour suppressor proteins. *Nature Cell Biol.* **5,** 578–581 (2003).
4. Inoki, K., Li, Y., Xu, T. & Guan, K. L. Rheb GTPase is a direct target of TSC2 GAP activity and regulates mTOR signaling. *Genes Dev.* **17,** 1829–1834 (2003).
5. Garami, A. *et al.* Insulin activation of Rheb, a mediator of mTOR/S6K/4E-BP signaling, is inhibited by TSC1 and 2. *Mol. Cell* **11,** 1457–1466 (2003).
6. Yarm, F. R. Plk phosphorylation regulates the microtubule-stabilizing protein TCTP. *Mol. Cell. Biol.* **22,** 6209–6221 (2002).
7. Liu, H., Peng, H.-W., Cheng, Y.-S., Yuan, H. S. & Yang-Yen, H.-F. Stabilization and enhancement of the antiapoptotic activity of Mcl-1 by TCTP. *Mol. Cell. Biol.* **25,** 3117–3126 (2005).
8. Yang, Y. *et al.* An N-terminal region of translationally controlled tumor protein is required for its antiapoptotic activity. *Oncogene* **24,** 4778–4788 (2005).
9. Thaw, P. *et al.* Structure of TCTP reveals unexpected relationship with guanine nucleotide-free chaperones. *Nature Struct. Biol.* **8,** 701–704 (2001).
10. Lee, Y. S. & Carthew, R. W. Making a better RNAi vector for *Drosophila*: use of intron spacers. *Methods* **30,** 322–329 (2003).
11. Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118,** 401–415 (1993).
12. Stowers, R. S. & Schwarz, T. L. A genetic method for generating *Drosophila* eyes composed exclusively of mitotic clones of a single genotype. *Genetics* **152,** 1631–1639 (1999).
13. Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* **22,** 451–461 (1999).
14. Patel, P. H. *et al.* *Drosophila* Rheb GTPase is required for cell cycle progression and cell growth. *J. Cell Sci.* **116,** 3601–3610 (2003).
15. Oldham, S., Montagne, J., Radimerski, T., Thomas, G. & Hafen, E. Genetic and biochemical characterization of dTOR, the *Drosophila* homolog of the target of rapamycin. *Genes Dev.* **14,** 2689–2694 (2000).
16. Zhang, H., Stallock, J. P., Ng, J. C., Reinhard, C. & Neufeld, T. P. Regulation of cellular growth by the *Drosophila* target of rapamycin dTOR. *Genes Dev.* **14,** 2712–2724 (2000).
17. Brogiolo, W. *et al.* An evolutionarily conserved function of the *Drosophila* insulin receptor and insulin-like peptides in growth control. *Curr. Biol.* **11,** 213–221 (2001).
18. Tapon, N., Ito, N., Dickson, B. J., Treisman, J. E. & Hariharan, I. K. The *Drosophila* tuberous sclerosis complex gene homologs restrict cell growth and cell proliferation. *Cell* **105,** 345–355 (2001).
19. Potter, C. J., Huang, H. & Xu, T. *Drosophila* Tsc1 functions with Tsc2 to antagonize insulin signaling in regulating cell growth, cell proliferation, and organ size. *Cell* **105,** 357–368 (2001).
20. Ito, N. & Rubin, G. M. gigas, a *Drosophila* homolog of tuberous sclerosis gene product-2, regulates the cell cycle. *Cell* **96,** 529–539 (1999).
21. Gao, X., Neufeld, T. P. & Pan, D. *Drosophila* PTEN regulates cell growth and proliferation through PI3K-dependent and -independent pathways. *Dev. Biol.* **221,** 404–418 (2000).
22. Goberdhan, D. C., Paricio, N., Goodman, E. C., Mlodzik, M. & Wilson, C. *Drosophila* tumor suppressor PTEN controls cell size and number by antagonizing the Chico/PI3-kinase signaling pathway. *Genes Dev.* **13,** 3244–3258 (1999).
23. Huang, H. *et al.* PTEN affects cell size, cell proliferation and apoptosis during *Drosophila* eye development. *Development* **126,** 5365–5372 (1999).
24. Tuynder, M. *et al.* Biological models and genes of tumor reversion: cellular reprogramming through tpt1/TCTP and SIAH-1. *Proc. Natl Acad. Sci. USA* **99,** 14976–14981 (2002).
25. Tuynder, M. *et al.* Translationally controlled tumor protein is a target of tumor reversion. *Proc. Natl Acad. Sci. USA* **101,** 15364–15369 (2004).
26. Walch-Solimena, C., Collins, R. N. & Novick, P. J. Sec2p mediates nucleotide exchange on Sec4p and is involved in polarized delivery of post-Golgi vesicles. *J. Cell Biol.* **137,** 1495–1509 (1997).
27. Guo, X. *et al.* A Rac/Cdc42-specific exchange factor, GEFT, induces cell proliferation, transformation, and migration. *J. Biol. Chem.* **278,** 13207–13215 (2003).
28. von Lintig, F. C., Pilz, R. B. & Boss, G. R. Quantitative determination of Rap 1 activation in cyclic nucleotide-treated HL-60 leukemic cells: lack of Rap 1 activation in variant cells. *Oncogene* **19,** 4029–4034 (2000).
29. Im, E. *et al.* Rheb is in a high activation state and inhibits B-Raf kinase in mammalian cells. *Oncogene* **21,** 6356–6365 (2002).
30. Sharma, P. M. *et al.* Inhibition of phosphatidylinositol 3-kinase activity by adenovirus-mediated gene transfer and its effect on insulin action. *J. Biol. Chem.* **273,** 18528–18537 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge H. Bellen, A. Bergmann, S. Cohen, B. Edgar, G. Halder, H. Richardson, G. Struhl, T. Xu, A. Selvaraj, G. Thomas and the Bloomington Stock Center for fly stocks, and the *Drosophila* Genomics Resource Center for cDNA clones. We thank H. Andrews, K.-O. Cho, G. Halder, J. Lim, S.-C. Nam, G. Roman and A. Singh for critical comments; R. Atkinson for assisting image analysis; and M. Acar for suggestions on S2 cell assays. We also thank G. Boss for advice on *in vivo* measurement of dRheb activation. Confocal microscopy was provided by an NIH core grant. This work was supported by NIH grants to M.L. and K.-W.C.

Author Contributions Y.-C.H. did most of the included studies; J.C. contributed to the initiation of this project and generated some dTCTP reagents, including antibody and dTCTP transgenic flies; Y.C and M.L performed *in vitro* GEF assays; K.-W.C. supervised the research project and data analysis.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.-W.C. (kchoi@bcm.tmc.edu).

naturejobs

THE CAREERS MAGAZINE FOR SCIENTISTS

Scientists bemoaning the lack of fresh talent entering some fields, such as maths and chemistry, sometimes look to popular culture for a boost. Portraying scientists at work in television and film can help to inspire and interest potential students. A good example of this effect is the rising popularity of crime shows such as *CSI: Crime Scene Investigation*, which has helped to encourage interest in forensic science.

But as Richard Smith notes in this week's Recruiters (see page 794), there can be a slight downside. The interest in forensic science has seen a deluge of applications for a handful of jobs. Perhaps more significantly, those applying are seeing the jobs through the eyes of fiction and are not really equipped with the skills and training needed to excel in their chosen profession. In Britain, one company is helping to launch better training programmes to ground aspiring scientists in the more mundane skills that off-camera researchers use in their day-to-day work.

With Will Smith set to hit the silver screen later this year as a virologist fighting vampires in the aftermath of biowarfare, maybe other sectors of science should brace themselves for a similar deluge of applicants. Smith reportedly spent some time researching his role for *I Am Legend* at the US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia. Perhaps he in turn will help boost interest in the CDC's biosafety role.

In much the same way as forensic science has seen both good and bad things result from its higher profile, *I Am Legend* could prove to be a double-edged sword. If the film gets more young people interested in science, that's a plus. In particular, there could be beneficial effects if it fuels interest in biosafety research, as this sector requires a long training period and currently has relatively few qualified people ready to work and an expanding number of facilities needing new recruits. But if the interest generated isn't matched by sufficient training and outreach programmes to help the newly curious find proper education and jobs, then it will be a missed opportunity.

Paul Smaglik, *Naturejobs* editor

CONTACTS
Editor: Paul Smaglik
Assistant Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
***Nature*events:**
Claudia Paulsen Young
(+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com

France/Switzerland/Belgium:
Muriel Lestringuez (4994)
UK/Ireland/Italy/RoW:
Nils Moeller (4953)
Scandinavia/Spain/Portugal:
Evelina Rubio-Morgan (4973)
Germany/Austria/The Netherlands:
Reya Silao (4970)
Online Job Postings:
Matthew Ward (+44 (0) 20 7014 4059)

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
***Naturejobs* web development:**
Tom Hancock
***Naturejobs* online production:**
Catherine Alexander

US Head Office, New York
75 Varick Street,
9th Floor,
New York,
NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

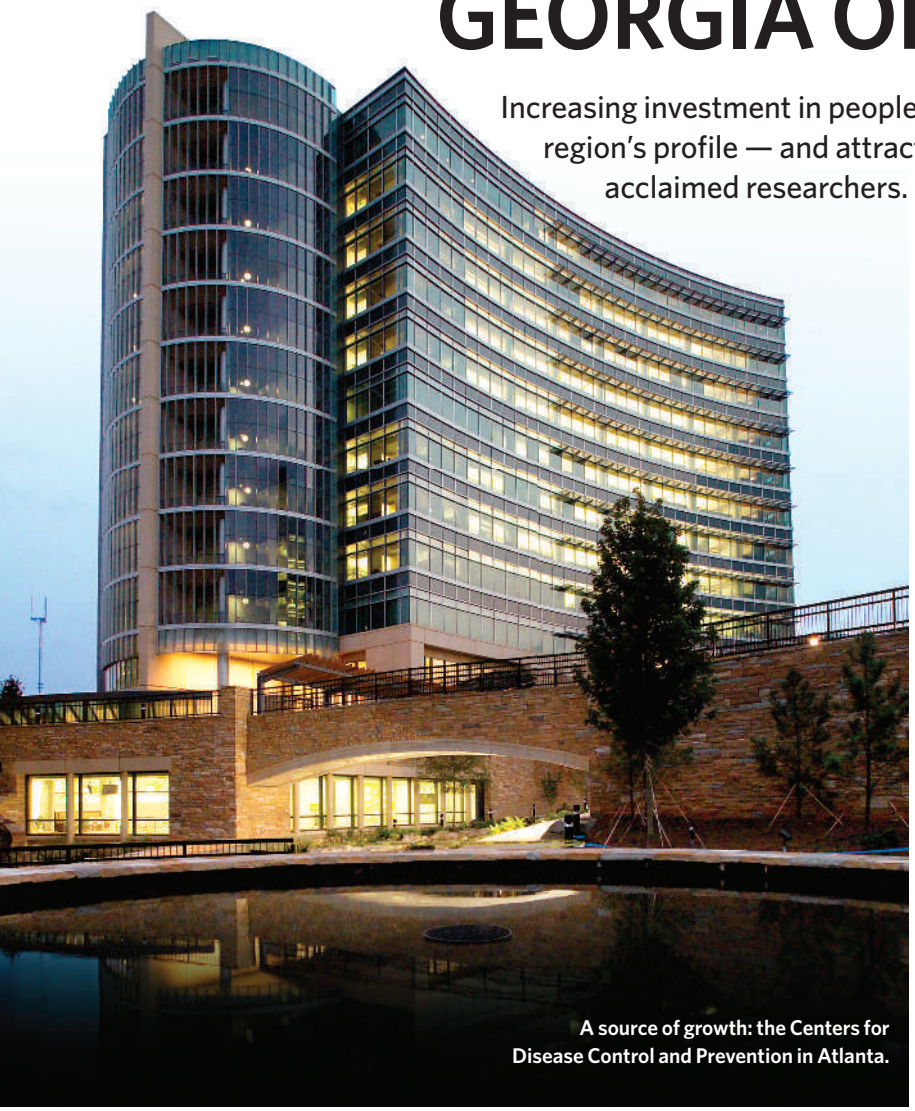
US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku,
Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
e-mail: a.watanabe@natureasia.com

GEORGIA ON THE MIND

Increasing investment in people is raising the Atlanta region's profile — and attracting a stream of internationally acclaimed researchers. **Paul Smaglik** reports.



A source of growth: the Centers for Disease Control and Prevention in Atlanta.

J. GATHANY/CDC

CNN and Coca-Cola are probably the best-known companies in Atlanta, Georgia. But investment in science by the state and federal governments and by research foundations is helping the city to gain scientific visibility. A programme to attract world-class scholars is at the heart of Georgia's rise in research rankings, international scientific recognition and job growth. It has brought in scientists who have been awarded individual grants, large federal projects, foundation funding and business involvement.

Georgia's research institutions began rising to prominence in 1990, when administrators from several of the state's universities banded together to form the Georgia Research Alliance (GRA). Since 1993, the group has invested more than $350 million in an effort to recruit 'eminent scholars', by offering salary top-ups and infrastructure. Those staff, in turn, have brought in existing grants and are better positioned to compete for additional funds. So far, 54 such renowned researchers are estimated to have brought $2 billion into Georgia's economy, says Mike Cassidy, president of the GRA.

The investment is paying dividends. Every year, consultancy firm Ernst & Young assesses US states in terms of the number of biotech companies that they have within their borders. For years, Georgia has languished outside the top ten, but last year it reached the number seven slot. Perhaps more impressively, the state has seen the amount of funding it attracts from the National Institutes of Health (NIH) more than double from $166 million in 1998 to $375 million in 2005. The single biggest recipient of this money is Emory University, which rose from number 31 in 1996 to 19 in 2005 in the NIH rankings of single institutions receiving grants.

As well as improving the rankings of Georgia research institutions, the GRA programme has had a knock-on effect on recruitment. The quality of Emory's seven 'eminent scholars' attracts graduate students, postdocs and junior professors, says Mary Delong, the university's director of postdoc education, who herself was attracted to Emory a month ago after working at the NIH in Bethesda, Maryland. "It's like a lightning rod for attracting capable young people," she says.

Attractive forces

One of Emory's high-profile names is Rafi Ahmed, director of the university's vaccine centre. Since he arrived from the University of California, Los Angeles, in 1995 to start the centre, he has recruited 22 faculty members and averaged $17.5 million a year in NIH funding. He has managed to draw in the money by winning a series of large programme grants in vaccine development, including several NIH-supported HIV vaccine and immunology projects and, most recently, a $4.5-million grant from the Bill & Melinda Gates Foundation for AIDS research. Ahmed says his interactive, collaborative approach helps in both grant-writing and research success. "We run a focused, coherent programme and we recruit coherently," he says.

The GRA recruited Julia Hilliard to Georgia State University in 1997. Apart from granting her eminent-scholar status, it built her a biosafety level-4 facility so that she could continue her work on pathogens that affect the central nervous system. "That was really appealing to me," says Hilliard, who supervises a team of 30. Georgia State University and the GRA shared the $1-million cost of building the facility, but Hilliard has received another $1 million a year from the NIH to run it.

> **"We are able to recruit high-quality talent as a result of the 'eminent scholar' programme."**
> — David Hartnett

Some of the area's most ambitious efforts go beyond single institutions. Donald Edwards, whose research spans physics and biology, helped obtain one of the largest single grants from the National Science Foundation ever awarded — $20 million to start the Center for Behavioral Neuroscience. The centre involves eight institutes, including Emory,

SWIMMING WITH SHARKS

The Georgia Aquarium is Atlanta's newest tourist attraction. Housing more than 100,000 animals in 8 million gallons of fresh and salt water, the site has been visited by more than 4 million people since it opened in November 2005.

A slice of every visitor's admission fee goes towards research targeted at conservation efforts. So far, the initiative has raised $500,000, says Bruce Carlson, vice-president for education, exhibits and conservation at the aquarium. "Our research is focused on our exhibits and animals we have here," he says. In particular, the aquarium's largest and most popular attractions are being studied: three whale sharks. "There's no textbook on the biology of whale sharks," Carlson notes.

The shark work extends beyond the aquarium's 50 or so professional staff — the centre is also funding researchers at the Mote Marine Laboratory in Sarasota, Florida, to study the sharks' habitat to help the Mexican government understand the role of environment and tourism on the sharks' population in the Gulf of Mexico. "By understanding these animals, we're hoping to help the Mexican government better manage them," Carlson says. The aquarium also funds a postdoc in Taiwan, who tags and monitors the sharks.

Closer to home, the aquarium offers veterinary students at the University of Georgia in Athens a great training opportunity. Student interns can get experience of treating animals at the aquarium. "There aren't a lot of opportunities for students in veterinary medicine to get hands-on aquatic medicine experience," says Carlson. As visitor numbers — and the revenue from them — keep growing, the aquarium intends to expand its research efforts and has already begun sponsoring a coral-reef research conservation effort for coral reefs of the Solomon Islands in the South Pacific. **P.S.**

RON CHAPPLE STOCK/CORBIS

The Georgia Aquarium is funding conservation research on the whale shark (inset).

Georgia State University and the Georgia Institute of Technology. Edwards is looking for multidisciplinary postdocs to help staff the facility.

In several cases, winning these large grants built upon earlier investments by universities and the GRA. That is the case for a nanomedicine centre set up by the Georgia Institute of Technology, Emory University and the Medical College of Georgia. The centre will receive between $6 million and $10 million from the NIH over the next five years, and almost $3 million from the GRA.

Strong partners

"It's a very successful collaboration between a medical school and engineering school using the strength of both, as well as a collaboration by a public school, a private school and a medical school," says Gary Schuster, provost at the Georgia Institute of Technology. Schuster says that the GRA has been successful at matching up each institute's strengths with others' needs — for example, having engineers at 'Georgia Tech' develop instrumentation for experiments at Emory or the Medical College of Georgia.

The GRA's eminent-scholar programme has also had an impact on Georgia-based businesses. David Hartnett, vice-president of technology industry expansion for the Metro Atlanta Chamber, and former chief executive of four area start-ups, sees the programme as a way to draw talent into the region's high-tech businesses. "We are able to recruit high-quality talent as a result of the programme," he says. BresaGen, an Australian company, chose Athens, Georgia, as its US headquarters in part to continue its work with eminent scholars Steven Stice and Clifton Baile at the University of Georgia. Several eminent scholars are also involved in starting and managing businesses. Nikil Jayant, for example, co-founded EGT, an Atlanta-based telecommunications software firm.

The biggest infrastructure development in Atlanta is at the Centers for Disease Control and Prevention (CDC). Since the terror attacks of 11 September 2001, the US government has pumped more than $1.5 billion into the centre to make it better prepared to handle bioterror threats. The Emerging Infectious Diseases Laboratory triples the CDC's capacity to research pathogens that require the highest levels of safety precautions, such as Ebola, viral haemorrhagic fevers, monkeypox and avian influenza. It includes new biosafety level-3 and level-4 facilities, which have been constructed but not yet commissioned.

But the new facilities don't necessarily mean a plethora of new positions, says CDC director Julie Louise Gerberding. The new facilities replace outdated ones, she says. Still, Gerberding says that the CDC's role in infectious disease ensures a steady stream of fellows and visiting scientists. In US epidemiology, "all roads lead to the CDC", she says. And with more programmes to attract world-class scientists to Georgia, many more roads are leading to Atlanta. ■

Paul Smaglik is editor of *Naturejobs*.

G. KNOBLOCH/CDC

Atlanta is seeing a rise in biosafety facilities and related work.

NATURE|Vol 445|15 February 2007

MOVERS

Bertil Andersson, Provost, Nanyang Technological University, Singapore

2004–07: Chief executive, European Science Foundation, Strasbourg, France
1999–2003: President, Linköping University, Linköping, Sweden
1996–99: Dean, chemical-sciences department and pro-dean of the science faculty, Stockholm University, Sweden

Nothing appeals to Bertil Andersson like speed. As an undergraduate, he was captivated by the fast biochemical reactions taking place in nerves. But his supervisor directed him towards reactions in photosynthesis — a field of growing interest in the mid-1970s.

Andersson pursued a graduate degree in biochemistry at the then newly established Umeå University in northern Sweden, which had a pioneer spirit — perfect to encourage his interest in the burgeoning field of plant molecular biology. As a postdoc at Australia's Commonwealth Scientific and Industrial Research Organisation, he helped decipher the now textbook descriptions of chlorophyll–protein complexes and their role in energy conversion.

He returned to Sweden as an associate professor in biochemistry at Lund University, and quickly became chair of Stockholm University's biochemistry department at 37 — a pivotal moment in his career. He discovered a talent for administration and was soon promoted to dean, which led to committee chair positions on the Swedish Biochemical Society and the Nobel Committee for Chemistry. "I made the decision to manage well and sleep less," he says.

But juggling research and administrative duties eventually became frustrating. Andersson accepted the post of chief executive at the European Science Foundation, to focus on creating a research arena that could better compete with the United States, Japan and rapidly advancing countries such as China. He's most proud of a grant programme that champions Europe's finest early-career scientists. He also designed a common procedure for the peer review of grant applications, which had been plagued by fragmented, nation-specific approaches.

Now he has been lured to a part of the world famous for its rapid innovation. In April, Andersson will become the first provost of Nanyang Technological University (NTU), a major university in Singapore boasting 25,000 students.

"We rightly believed that strong scientists, such as Andersson, would be attracted to a system that has ready funding and few boundaries to create research programmes," says Haresh Shah, Stanford University professor emeritus in engineering and member of the NTU's board of trustees.

Andersson says he wants to take advantage of how quickly the Singaporean government moves from idea to implementation — and he's excited about connecting his European and US colleagues with Singapore's opportunities and abundant science resources. ■

Virginia Gewin

NETWORKS & SUPPORT

Goodbye Poland, again

In 1999, I left my home country of Poland, where I had earned an MD and a PhD in neuroscience, to pursue postdoctoral training in the United States. After five years I returned to my homeland, hoping to set up my own lab there. It wasn't to be.

I was lucky enough to obtain a short-term contract, courtesy of my PhD supervisor. But I was on a long waiting list, with little chance of gaining research independence. After a year, I returned to the United States, where I took a junior faculty position at Johns Hopkins University.

Poland loses biomedical talent because its academia has difficulty accommodating new independent laboratories. A few institutions use outside recruits to replace retired scientists. Most, though, engage in 'inbreeding', which contributes to an inert and defective system. Internal recruitment also puts candidates who spend years abroad at a disadvantage. Young researchers determined to stay in their homeland often take dead-end technical positions or wait years for a chance to set up a lab.

The West offers financial support in the form of programmes to help Polish scientists return, such as the international researcher programmes at the Howard Hughes Medical Institute or the Wellcome Trust. But these are more likely to help local scientists who have already been accepted by home institutions and supported by local foundations. They do little for 'excess' scientists because they don't create new laboratory space.

In a European country with a long communist history, only government can help academia. But biomedical research and development issues are not on any political party's agenda. The ruling parties, obsessed with historic and nationalistic issues, have little interest in innovation. Politicians do not see R&D as a good investment that facilitates development.

The European Union and philanthropists may help more by not pouring resources into academia but helping to develop the commercial side of science. This could be achieved, for example, by creating more technology-transfer offices at major universities and training Polish scientists in the West to work in them. With luck, this would help create the right climate for private investors, without whom the sustainability of the industry is hard to imagine.

Despite the unfriendly climate, a handful of medical biotech start-ups has appeared in recent years. A spirit of entrepreneurship has somehow survived. I hope it will be nurtured. ■

Arkadiusz Szklarczyk is a research associate at the Department of Neurology, Johns Hopkins University Medical School in Baltimore, Maryland, USA.

POSTDOC JOURNAL

Taking a gamble

I'm an ex-PhD student. It's been 15 months since I submitted my thesis, on the tectonics of New Zealand, and 10 months since I successfully defended it.

Where to go after your PhD? If you've got through without every spark of interest being crushed out of you (I did), and you still feel masochistic enough to brave further exposure to the academic world (I'm a glutton for punishment), a postdoc is the usual next step. But my step was sideways, into a rather split-personality technician/teaching post in my old lab at the University of Southampton, UK. Despite a secure job, and valuable teaching experience, the lack of opportunities for new research made me worry for the future of my CV.

So I find myself taking a bit of a gamble. This month I'm leaving England to take up a postdoc position in Johannesburg. I'll not only be living and working in a country I've never visited before, but I'll also be trying to work out the tectonic history of rocks almost two billion years older than any I've studied before. If that's not enough, I also need to determine whether I can turn some half-formed scientific ideas and interests into a coherent research programme. Do I have what it takes to stay in the game long term? Is that what I want? ■

Chris Rowan is about to become a postdoctoral fellow in the geology department at the University of Johannesburg, South Africa.

The inside track from academia and industry

Crime scene investigators

Forensic science is swamped with applicants, but a pure science degree will put you ahead of the crowd.

Richard Smith

For many people, forensics makes science sexy. A combination of intrigue, curiosity and élitism has captured the imagination of millions who regularly tune into programmes such as *CSI: Crime Scene Investigation*, *Waking the Dead* and *Silent Witness*. But forensics is not only popular material for television. Interest in the subject as a career has grown, reflected in the number of applicants for jobs in the sector.

In the decade or so that this kind of show has been popular, LGC Forensics, the largest independent forensic service in the United Kingdom, has seen a steep increase in the number of job applications. Now the company, which offers consultative and analytical forensics services to UK police forces and other agencies, routinely receives several thousand applications each year for as few as 50 vacancies. The growth seems to suggest a direct correlation with TV exposure, although the link is hard to prove. Before the explosion of forensic science in the media, there was a steady trickle of applications, usually in the hundreds.

As with most forms of recruitment, it is important to attract interest. But with thousands continuing to apply it can be a logistical nightmare trying to put together a shortlist for interview.

The university system in Britain also views forensic science as an opportunity to attract students. Hundreds of courses include 'forensics' in their titles — it's a way of selling science. As former chairman of a government-sponsored national enquiry into UK higher education and forensic science, I found that the subject attracted students who otherwise might not have pursued a science degree. This, one could argue, is a success story in its own right.

"A proposed master's degree will cover key areas, such as lab work and the legal process."

But some students and prospective job applicants mistakenly feel that pursuing a forensic-science degree will provide a major advantage over pure science qualifications. Angela Gallop, director of LGC Forensics, always encourages students to take a first degree in pure science, arguing that a grounding in science is the best preparation for a career in forensics or any analytical science area.

LGC Forensics' experience suggests that many (although not all) of the applicants who apply for jobs actually lack the 'pure science' skills and knowledge needed. Despite the increasingly large applicant pool, in some cases there was a need to provide several weeks of remedial training in areas such as analytical chemistry.

"It can take many years to gain the experience and skills needed to work at senior level."

Much of LGC Forensics' recruitment takes place at entry/junior level — developing skills in areas such as forensic biology, chemistry, toxicology and ecology. It can take many years of personal and professional development to gain the experience and skills needed to work at senior levels, which usually involves presenting forensic evidence in court and taking on case management. LGC Forensics has clearly defined and flexible development pathways, such as training in areas of forensic biology, followed by additional modules that will let employees build on their knowledge and skills.

Looking to foster innovative approaches as the demand for forensic services continues to grow, LGC approached the University of Oxford to explore a new idea — the development of an MSc in applied analytical science. This would cover key areas of forensic and analytical practice, including investigations at crime scenes, forensic science in the laboratory, and legal process in courts of law. The proposed degree builds on existing master's degrees in forensic science, for example at the University of Strathclyde in Scotland and King's College London, to accommodate the skills needed within the company. It is also likely to benefit other businesses and employers in the area of forensic–analytical science.

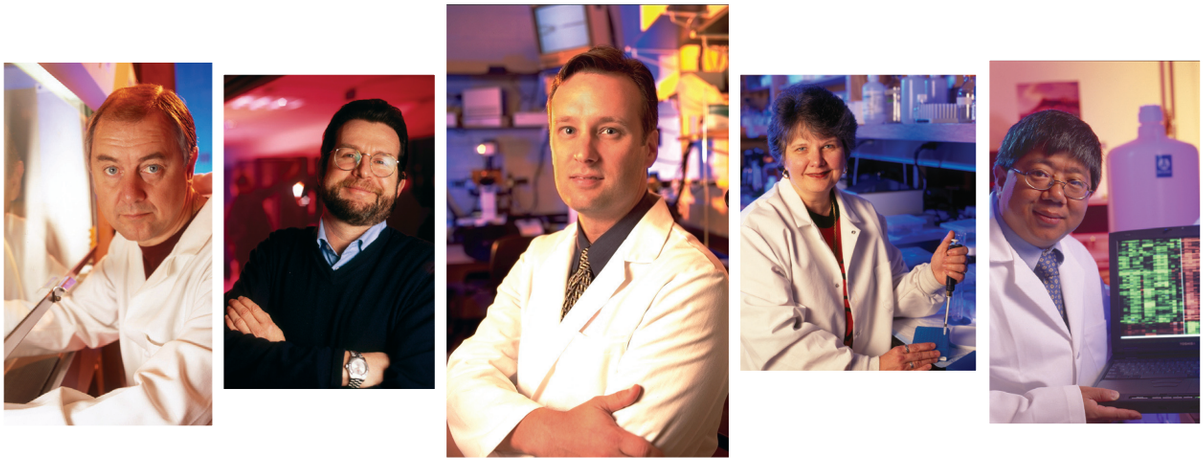
LGC Forensics has developed an outline programme composed of science subjects, the legal process and investigation. It has also made a commitment to provide expertise on curriculum development and teaching, including work placements. This approach is aligned with the UK government's call for industry–academic collaborations that offer greater cohesion and mutual benefit.

Our Oxford collaborators believe that this programme is an opportunity to provide a world-class course in vital areas of analytical science. For LGC, it will provide a valuable development tool for people in the company, while enabling other students to get a firmer grasp of the realities of forensics and analytical science. We hope to start accepting students for the 2009 academic year.

Through programmes such as these, we can develop the analytical skills of the future and ensure that forensics continues to be populated by specialists who possess both an intense curiosity and a knack for science. ■

Richard Smith is project director for the LGC Forensics/Oxford Programme and associate director of Winchester Consulting in Hampshire, UK.

- www.lgc.co.uk
- www.wincon.co.uk



Go where the talent is.

Georgia™

Tap into a world of possibility led by the Georgia Research Alliance Eminent Scholars.® These enterprising scientists not only make breakthrough discoveries in the laboratory, they work closely with the industry to help scientific and medical companies of all sizes flourish. With world-class research facilities and six nationally recognized universities. Georgia's resources and funding are the catalyst you need to realize your dreams. To find out more, contact us at 404-962-4006. Visit **www.GeorgiaAllies.com**. **Put your dreams in motion.**

NW96245R

The University of Georgia

Not Your Usual Power Plant

Photo courtesy of USDA NRCS

At the University of Georgia, bioenergy research is on fast forward.

Science and engineering talent at the University of Georgia combine to create a national capability to address energy, environmental and sustainability issues facing the state, the nation and the world. UGA, the nation's first land-grant university and the state's largest and most comprehensive public university, will play a key role in the State of Georgia's new bioenergy initiative. Our goal is to unite energy security and sustainable forestry and agriculture in ways that will produce solutions for rural economic development and global warming. UGA's bioenergy program builds on the University's strengths, including:

- Crop sciences
- Agribusiness
- Plant genomics
- Glycobiology
- Bioengineering
- Structural biology
- Microbiology
- Forest biotechnology
- Biochemistry

Incredible Anaerobes: From Physiology to Genomics to Fuels

A 'State-of-the-Art' Symposium, March 2 - 3, 2007
A two-day conference with 23 plenary talks and a poster session. Leading researchers will evaluate the science of anaerobic microorganisms in biofuel production from renewable biomass.

- Anaerobes for Biomass Conversion
- Cellulose Degradation
- Biomass to Fuel Conversion

Georgia Center for Continuing Education, University of Georgia, Athens, GA
www.georgiacenter.uga.edu/conferences/2007/Mar/02/anaerobes.phtml

The University of Georgia is an Affirmative Action/Equal Opportunity Employer. www.uga.edu.

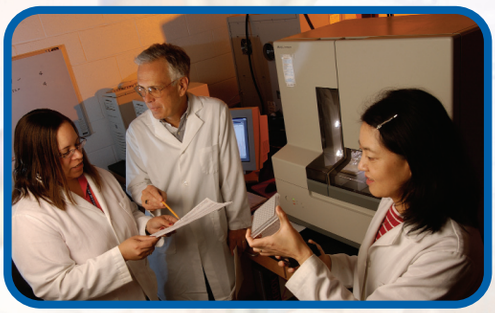
The Georgia Research Alliance (GRA) Endowed Chair in Bioenergy — UGA invites applications from accomplished scientists for an endowed chair in bioenergy research. For information, contact Professor Michael W. W. Adams (Chair), GRA Eminent Scholar in Bioenergy Search Committee, adams@bmb.uga.edu.

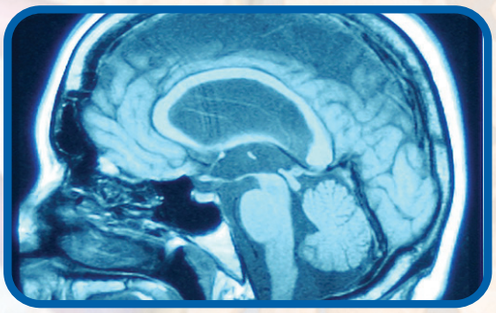
NW96751A

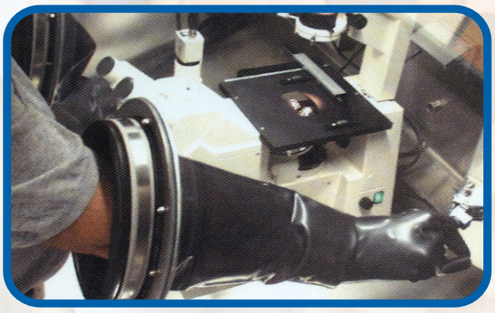
Georgia State University

DISCOVERY & DEVELOPMENT

Georgia State University, in the heart of downtown Atlanta, provides the cutting-edge research, top-quality facilities and highly educated, diverse workforce that the science-based industries need to succeed. The university has created several core initiatives that have allowed Georgia State University and the state of Georgia to stay at the forefront of bioscience and neuroscience research and education:







- Interdisciplinary areas of focus in Molecular Basis of Disease, Brains and Behavior, and Urban Health.
- A new Science Park, currently under construction, which will provide unique research and learning opportunities for our faculty, our students, and for industry.
- An internationally-known faculty, including a number of Georgia Research Alliance Eminent Scholars, each recognized as a driving force in his or her field.
- Clinical diagnostic tests and research conducted in state-of-the-art BSL 3 and 4 maximum containment facilities.
- Alliances for building new biotechnology firms and projects, including core research facilities and incubator space for startups.
- Collaborative research and graduate education programs in the biological, chemical, computational, and behavioral sciences.

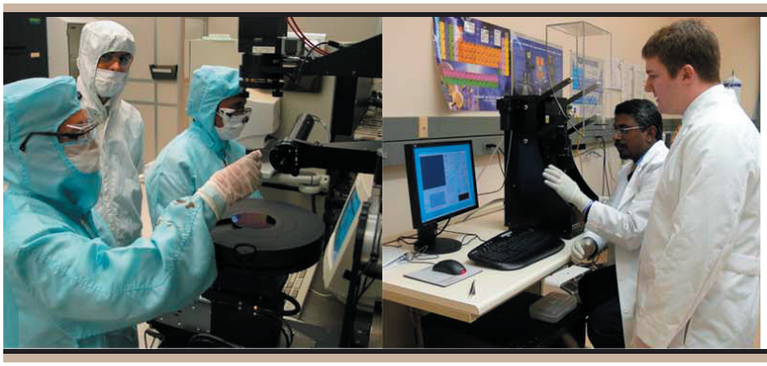
FOR MORE INFORMATION:

Industry and Commercial Contacts
Mr. Joe Gilbert, University Research Services and Administration, jgilbert@gsu.edu (404-463-4745)

Research and Academic Contacts
Dr. William Nelson, College of Arts and Sciences Dean's Office, wnelson@gsu.edu (404-651-2294)

WWW.GSU.EDU

NW96586R



Microelectronics Research Center

Georgia Institute of Technology

The Georgia Institute of Technology's Microelectronics Research Center (MiRC) provides expertise, facilities, infrastructure and teaming environments to enable and facilitate interdisciplinary research in microelectronics, nanotechnology, integrated optoelectronics, process development, semiconductor physics, biomedical devices and applications, as well as microsensors, microfluidics and actuators. Our users include traditional academic researchers, as well as industry, government agencies, and start-up companies. As the southeastern node of the NSF sponsored National Nanotechnology Infrastructure Network, a thirteen institution strong network of facilities across the U.S., we also serve as the portal to a national network of nanoscience facilities and technical resources. Our mission supports a wide variety of research programs in nanotechnology with a special focus on Biomedical Applications, and Educational Outreach.

Our MiRC facility is housed in a 100,000 square foot building and a 20,000 square foot annex and provides facilities which include six electronic and optoelectronic materials labs, eight labs for microelectronics design and testing, eight labs for electronic device design and testing, and a 85,000 square foot cleanroom providing comprhensive microfabrication and nanofabrication facilities. Among its nanofabrication tools the MiRC has a state of the art electron beam lithography system that has demonstrated sub-7 nm lines and sub 3.5nm gaps. This tool, a JEOL JBX9300FS, has 100 keV accelerating voltage and a 4 m spot size and is capable of writing on samples from ranging ~ 1 mm^2 to 300 mm wafers. The MiRC allows outside users access to this exceptional tool whether they come to the MiRC to be trained to operate it themselves or by submitting remote jobs that will be carried out by MiRC staff. Integrative to our technical facilities is an administrative area that includes offices and seating for facility staff, academic faculty, visiting researchers, and graduate students as well as meeting and interactive space for workshops, conferences, and presentations.

In Fall of 2006, Georgia Tech broke ground for our new Marcus Nanotechnology Building (NRB). Scheduled for completion in Fall 2008, this new 190,000 square foot building will have a 30,000 square foot ballroom style cleanroom. Within the cleanroom envelope 20,000 square feet will be dedicated to nanotechnology focused on the physical sciences and engineering, and the remaining 10,000 square feet of cleanroom will be dedicated to biological and biomedical nanotechnology research. Researchers are encouraged to carry out their work using the capabilities of both portions of the cleanroom. The NRB will supply the level of expertise, facilities, infrastructure, and teaming environments to enable interdisciplinary research in nanotechnology. Under the direction of Dr. James Meindl, 2006 IEEE Medal of Honor recipient, the MiRC and the Marcus Nanotechnology Building will lead in efforts to fuse multiple scientific disciplines in pursuit of breakthrough nanotechnologies.

Microelectronics Research Center

Marcus Nanotechnology Building, construction underway

Microelectronics Research Center
791 Atlantic Drive NW
Atlanta, GA 30332
404.894.5100

James D. Meindl, Ph.D.
Director, Microelectronics Research Center
Director, Marcus Nanotechnology Building

www.mirc.gatech.edu
grover.mirc.gatech.edu

Kevin Martin, Ph.D.
Associate Director
kevin.martin@mirc.gatech.edu
404.894.5030

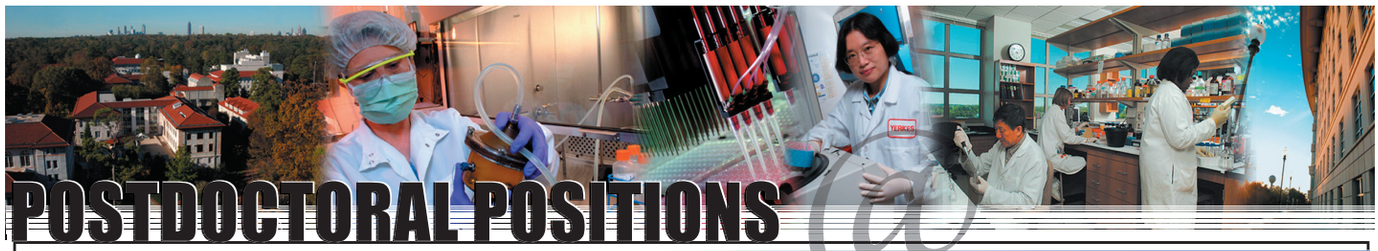
Tina Prestridge
Asst Director, Business Operations
tina.prestridge@mirc.gatech.edu
404.894.9431

Gregory Book, Ph.D.
Senior Research Engineer
gregory.book@mirc.gatech.edu
404.385.6963

Paul Joseph, Ph.D.
Research Scientist II
paul.joseph@mirc.gatech.edu
404.894.5029

Gary Spinner
Cleanroom Manager
gary.spinner@mirc.gatech.edu
404.385.4010

NW96753A



POSTDOCTORAL POSITIONS @

EMORY UNIVERSITY SCHOOL OF MEDICINE

Emory University School of Medicine in Atlanta GA invites talented researchers for postdoctoral positions available in Neuroscience, Cancer research, Pharmacology, Biochemistry, Immunology, Genetics, Pathology, Cardiology, Physiology, Microbiology, and diverse other areas. Each year 100 to 150 new postdoctoral fellows come to Emory School of Medicine.

In 2006, Emory University was named the top-ranked university and the number four institution overall in the "Best Places to Work for Postdocs" in a survey conducted by The Scientist magazine. Emory is among the top 20 US medical schools for Federal research funding and the top 5 for NIH NRSA fellowships.

Emory provides a well-funded, collaborative, state-of-the-art research environment enriched by nationally renowned clinical programs and interactions with adjoining CDC, multiple universities, and a growing biotech industry for interactive training, research and

EMORY UNIVERSITY SCHOOL OF MEDICINE

employment opportunities. Atlanta offers a moderate cost of living, great climate and diverse recreation/entertainment options. Competitive salaries, excellent full employee benefits including retirement benefits are available to all postdoctoral fellows. The Office of Postdoctoral Education offers numerous career development events and services.

The special Cottrell Postdoctoral Fellows Program provides $50,000 support/yr to outstanding applicants and the FIRST Postdoctoral Fellows Program provides special teaching training with research training to interested individuals.

See the Website below for a listing of Emory School of Medicine Departments and faculty and links of faculty research pages related to your research interests. Interested individuals should communicate directly with the faculty.

http://www.emory.edu/WHSC/MED/POSTDOC/

Office of Postdoctoral Education
Emory University School of Medicine

An Equal Opportunity/Affirmative Action Employer

NW96205R

A NATIONAL SYMPOSIUM:

PREDICTIVE HEALTH AND SOCIETY

December 17-18, 2007

Emory Conference Center, Atlanta, Georgia

Sponsored by the
Emory/Georgia Tech Predictive Health Initiative

For more information contact
Jennifer Vazquez at 404-712-2660,
Jennifer.vazquez@emory.edu

EMORY UNIVERSITY

Georgia Institute of Technology

NW93778E

EMORY VACCINE CENTER | EMORY UNIVERSITY

Where Science Meets Hope.

TENURE-TRACK FACULTY POSITION

The Division of Infectious Diseases of the Department of Medicine in conjunction with the Emory Vaccine Center invites application for tenure-track faculty positions (all ranks available commensurate with the applicant's record and experience) available immediately in the field of bacterial pathogenesis with a focus on tuberculosis research and vaccine development. Candidates should have interest in developing a program in the genetics, pathogenesis, immunology or prophylaxis of this important and worldwide chronic disease. A research interest in tuberculosis and co-infection with the human immunodeficiency virus (HIV) is also a consideration. Emory University School of Medicine and the Emory Vaccine Center have committed substantial resources and laboratory space to expand their programs in persistent infections, infectious disease pathogenesis and vaccine development with a mandate to expand an existing cadre of accomplished interactive scientists. The candidate should possess a PhD, MD, MD/PhD or DVM and have a strong background and successful publication record, as well as the potential to develop an active independent research program and to generate independent extramural funding. The successful candidate will be expected to form close and substantive ties with other faculty members in both the basic science and clinical departments.

Emory University is an equal employment opportunity/affirmative action employer. Women and minority candidates are strongly encouraged to apply. Interested candidates should send a letter of inquiry and at least three letters of reference by regular mail or e-mail to:

Edward S. Mocarski, Chair, Search Committee
Emory University School of Medicine
Suite 429, 1462 Clifton Rd., Atlanta, GA 30322

office: 404-727-9442 fax: 404-712-9736 e-mail: mocarski@emory.edu

NW96192R

The Medical College of Georgia Vascular Biology Center is recruiting a

Pulmonary Vascular Biologist

at the Assistant, Associate or Full Professor Level, Tenure-Track.

The successful candidate will have an earned Ph.D., M.D. or M.D./Ph.D. degree. He/she will join an active group of extramurally funded vascular biologists (currently about $8 million annually, see: http://www.mcg.edu/centers/VBC/index.html) in recently renovated laboratories utilizing state of the art equipment. He/she will have the opportunity to participate in the two institutional pre- and post-doctoral training programs in Integrative Cardiovascular Biology. Ample opportunities for collaborative basic and clinical research are available and encouraged. The candidate is expected to have and further develop an active, extramurally funded research program in aspects of pulmonary vascular disease, especially acute lung injury. Highly competitive salary and start-up package, commensurate with prior experience, will be provided. Applications should include detailed CV, statement of career goals and names of three references and e-mailed to John D. Catravas, Ph.D. (jcatrava@mcg.edu). The Medical College of Georgia is an AA/EOE. Applications from women and under-represented minorities are particularly encouraged.

NW96022R

Morehouse College
Assistant Professor in Chemistry

Tenure track position available beginning August 2007. Teaching responsibilities will include advanced inorganic chemistry and analytical chemistry.

A Ph.D. in analytical chemistry or a Ph.D. in inorganic chemistry, with specialization in materials science, is required. Post-doctoral and teaching experience are preferred.

The successful applicant must be able to establish an independent research program.

To apply, send a letter of application including curriculum vitae, transcripts, a state of teaching philosophy and a research plan, and arrange three letters of recommendation to be sent to: **Chemistry Search Committee, Department of Chemistry, Morehouse College, 830 Westview Drive, Attn: Dr. John H. Hall, Atlanta, GA 30314.**

Review of applications will begin on Monday, March 5th, 2007, and will continue until the position is filled.

Morehouse College is an EEO/AA employer and welcomes applications from diverse candidates.

NW94324R

NIH physician specialty training

If your goal is to pursue academic medicine, and you have a U.S. medical license, consider fellowship or residency training with the physicians who are performing basic research and designing pivotal trials that will determine state-of-the-art care for the next decade. The National Institutes of Health is home to 250 research beds. We offer outstanding clinical training and a research experience that includes intense exposure to designing and analyzing clinical trials, and to cutting-edge basic science investigation, within our ACGME-accredited programs.

Student Loan Repayment of up to $35,000 per year may be available to qualified individuals accepted into ACGME-accredited training programs.

For more information on ACGME-accredited programs and the Loan Repayment Program, visit our web sites at *www.training.nih.gov, www.lrp.nih.gov*

Office of Intramural Training and Education
Bethesda, Maryland 20892-1158
(888) 695-5343

NIH is dedicated to building a diverse community in its training and employment programs.

subspecialty training

Internal Medicine
Allergy and Immunology
Critical Care Medicine
Endocrinology and Metabolism
Hematology
Infectious Diseases
Medical Oncology
Rheumatology

Pediatrics
Allergy and Immunology
Endocrinology
Hematology/Oncology
Pediatrics/Medical Genetics

Blood Banking/Transfusion Medicine
Cytopathology
Hematopathology
Medical Genetics

residency training

Anatomic Pathology
Dermatology *(third year only)*
Psychiatry *(fourth year only)*

resident and medical student electives

Clinical and research electives for residents.
Flexible duration, 4 weeks, 8 weeks, up to one year.

The NIH Director's Wednesday Afternoon Lecture Series

Biomedical scientists around the world are invited to join us online to hear leading investigators present their latest results to the NIH Intramural Research community. Lectures may be viewed live at 3:00 p.m., EST (20:00 GMT) on Wednesdays, from September through June. Live webcasts can be viewed under "Today's Events" at: <http://videocast.nih.gov/>

The current schedule of lectures is available at: <http://www1.od.nih.gov/wals/schedule.htm>

Upcoming Lectures:

•February 21: Marc G. Caron, Duke University Medical Center: Novel GPCR Signaling Paradigms in Animal Models

•February 28: L. Mahadevan, Harvard University: Mathematics, Mechanics and Mobility

•March 7: Richard Flavell, Yale University of Medicine: Chromosome Dynamics in Cytokine Gene Expression

•March 14: Max Cooper, University of Alabama at Birmingham: Evolution of Two Adaptive Immune Systems

•March 21: Rick Morimoto, Northwestern University: Protein Misfolding in Aging and Neurodegenerative Disease

Tenure/Tenure-Track Position
Laboratory of Persistent Viral Diseases
Rocky Mountain Laboratories, Hamilton, Montana

National Institute of Allergy and Infectious Diseases (NIAID)
National Institutes of Health

The Laboratory of Persistent Viral Diseases (LPVD), Rocky Mountain Laboratories, NIAID, NIH, DHHS, in Hamilton, Montana, seeks applicants for a tenured or tenure-track position (full to assistant professor equivalent) to conduct independent research on host immune or inflammatory responses in neuropathogenic viral diseases. Candidates with a background in adaptive or innate immunity, including neuroinflammation and gliosis are preferred; those interested in neurobiology, biochemistry or pathogenesis of CNS infections are also encouraged to apply. Candidates must hold a Ph.D., D.V.M, or M.D. degree and have a minimum of 3 years of relevant postdoctoral experience. Candidates must be able to develop an independent research program, supervise staff and fellows, and collaborate with other LPVD researchers working on CNS viral or prion diseases.

Rocky Mountain Laboratories' state-of-the-art facilities include an operational BSL-3 facility, a BSL-4 lab and animal facility nearing completion, and in-house core facilities for genomics, electron microscopy, and flow cytometry. Research programs focus on prions, murine retroviruses, HIV, flaviviruses, and numerous pathogenic prokaryotic organisms. The lab is located in the scenic Bitterroot Valley of western Montana with easy access to some of the finest outdoor recreational opportunities in North America. Additional information on the position may be obtained by contacting Dr. Bruce Chesebro at bchesebro@niaid.nih.gov.

Application Process: Salary depends on degree and qualifications. To apply, submit a curriculum vitae and bibliography, including a list of your five most significant papers, and a 2-3-page description of a proposed research program, via e-mail to **Ms. Felicia Braunstein at braunsteinf@niaid.nih.gov**. In addition, three letters of recommendation must be sent directly from the referees to **Ms. Felicia Braunstein, Committee Manager, NIAID/NIH; 10 Center Drive, Bldg. 10, Rm.4A31, MSC-1356; Bethesda, MD 20892-1356**. Applications must reference **AD #010** and must be received by **March 9, 2007**. Applicants will be notified when their applications are received and then complete. All information provided by applicants will remain confidential.

Tenure Track/Tenure Investigator Positions in Systems Immunology and Infectious Disease Modeling

The National Institute of Allergy and Infectious Diseases (NIAID), Division of Intramural Research (DIR) is seeking several outstanding individuals for its new Program in Systems Immunology and Infectious Disease Modeling (PSIIM).

Modern technology allows the analysis of immune responses and host-pathogen interactions at multiple levels - from intracellular signaling networks, to individual cell behavior, to the functioning of a tissue, organ, and even the whole organism. The challenge is not only to collect large amounts of data, but also to organize it in a manner that enhances our understanding of *how* the immune system operates or *how* pathogens affect their hosts. To do this, we need to develop detailed quantitative models that can be used to predict the behavior of a complex biological system. These models can help to explain the mechanisms underlying physiological and pathological responses to infection or vaccination, which can then be exploited to design better therapies or vaccines.

Achieving this goal requires an interdisciplinary effort and to this end the PSIIM will be organized as an integrated team of scientists and support staff with expertise in computational biology, bioinformatics, proteomics, cell biology, immunology, and infectious diseases, rather than as a group of independent laboratories. These teams will have access to the latest technology for gene-expression profiling, high-content screening of RNAi libraries for the discovery of pathway components, imaging tools, cores for the genetic manipulation of animals and for proteomic analysis, and a substantial computer infrastructure. BSL-3 facilities for working with high priority pathogens will also be available.

The PSIIM is now recruiting for tenure-track or tenure level team leader appointments in three key areas:

Computational Biology: The incumbent will lead a group focused on the development and improvement of software tools for multiscale modeling and simulation that can be used by the PSIIM as well as by biologists interested in subjects other than immunity or infectious diseases. The ideal candidate will have a strong background in mathematics, physics, and computer programming, and a clear desire and ability to interact with and support the efforts of biologists. A demonstrated ability to generate computer software tools for biological modeling will be a strong plus.

Molecular/Cell Biology: The incumbent will lead a group involved in the design, implementation, and interpretation of screening efforts to identify and determine the interactions among the components in signaling networks that could then be modeled using the software generated by the computational biology team or obtained from other sources. Discovery tools such as gene arrays, high content image–based screens using RNAi methods, various protein-protein hybrid screening methodologies, and optical imaging are expected to be key elements in the efforts of this group. A strong background in basic cell biology and molecular biology with experience in analysis of protein-protein interactions, signaling, and/or gene regulation is required. Expertise in large-scale screening is highly desirable.

Infectious Diseases: The incumbent will be responsible for developing novel approaches to systems-wide analysis of the interaction of infectious agents and their hosts. These may include the use of gene expression signatures, the production of gene modified animals, the development of methods for *in vivo* testing of the predictions of models, and the use of sophisticated imaging and other tools for probing the interaction of pathogens and host cells *in vitro*. A strong background in viral and/or bacterial infectious diseases and cell and molecular biology are necessary; training in the immunology of infectious diseases and substantial bioinformatics experience are highly desirable.

These positions and the research activities they conduc are fully funded by the intramural research program of NIH. Each team leader is expected to build a working group consisting of postdoctoral fellows, staff scientists, technicians, and students. The team leaders will work with the program director to help set the goals for the PSIIM and to determine how best to reach these goals as an integrated group. To ensure appropriate career trajectories for those joining the PSIIM team, the NIH has modified its tenure policies to encourage and account for contributions made in such a team science setting. Applicants should be seeking a difficult challenge in which creativity, technical expertise, and a strong desire to achieve in a team environment are critical for success.

Interested candidates may contact **Dr. Ronald Germain, Program Director, PSIIM, DIR, NIAID at 301/496-1904 or email (rgermain@niaid.nih.gov)** for additional information about these positions.

To apply, submit your curriculum vitae, bibliography, and detailed statement of how you can contribute to the success of the PSIIM program to: **Felicia Braunstein at braunsteinf@niaid.nih.gov**. In addition, three letters of reference must be sent directly from the referee to **Dr. Robert Hohman, Chair, NIAID Search Committee, c/o Ms. Felicia Braunstein, DIR Committee Management Team Lead, 10 Center Drive, MSC 1356, Building 10, Room 4A31, Bethesda, Maryland 20892-1356.** Completed applications **MUST** be received by **February 16, 2007 for computational biology, March 16, 2007 for molecular/cell biology, as well as for infectious diseases. Please refer to ad #012 for computational biology, #013 for Molecular/Cell Biology and #014** for **infectious disease** on all correspondence. Further information on these positions and guidance on submitting your application are available at: **http://healthresearch.niaid.nih.gov**. For more information about NIAID systems biology program, please visit **http://www.nih.gov/catalyst/2006/06.09.01/page1.html**

NW96730R

OPPORTUNITIES @ NIH THE NATIONAL INSTITUTES OF HEALTH